REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD Year 2005 Name of Author KIRK, N. S

Structural and functional analysis of the proteasome inhibitor PI31 and its interaction with the F-box protein Fbxo7

by

Rebecca Jane Kirk

A thesis submitted to the University of London for the degree of

Doctor of Philosophy

2005

CANCER RESEARCH UK

London Research Institute
Lincoln's Inn Fields
London
WC2A 3PX

UCL

University College London
Gower Street
London
WC1E 6BT

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

This thesis describes the structural and functional analysis of two related proteins, PI31 and Fbxo7, both of which act within the ubiquitin-proteasome system. This system is involved in a wide variety of cellular processes. Proteins modified with ubiquitin are often targeted for degradation by the proteasome, an ATP-dependent multi-subunit complex. Specificity for ubiquitin transfer is controlled by the E3 ubiquitin ligases. The multi-subunit SCF E3 ubiquitin ligases use their F-box domain protein to engage substrates.

Analysis of possible binding partners of the novel F-box protein Fbxo7 using an affinity pull-down in Jurkat cell lysates identified PI31, a putative proteasome inhibitor. Fbxo7 was also shown to bind Skp1, Cullin1 and Rbx1 inferring that it forms part of a functional E3 ubiquitin ligase. Crystallisation of the Fbxo7-Skp1 complex yielded poorly diffracting crystals with a maximum resolution of 8Å.

The PI31 N-terminal domain (NTD) shows significant sequence similarity to a globular domain of Fbxo7. The PI31-NTD structure was obtained at 2.64Å by MAD phasing after introducing an additional methionine to engineer a non-centrosymmetric selenium substructure. The structure reveals a homodimer of novel α/β topology, which we define as the FP (Fbxo7 – PI31) domain since its distribution is limited to Fbxo7 and PI31 proteins in higher eukaryotes.

Biophysical analysis of structure-based mutations identified the relevant residues in the homodimer interface by confirming the production of a monomeric form of PI31. Further mutations also revealed the binding surface between Fbxo7 and PI31 by using isothermal calorimetry.

Fbxo7 and PI31 were shown to interact *in vivo* by co-immunoprecipitation and yeast two-hybrid screens using either PI31 or Fbxo7 as bait. Fbxo7 was shown to ubiquitinate FLAG-tagged PI31 *in vivo*. Abolishing the interaction between Fbxo7 and PI31 by mutations in PI31 prevents this ubiquitination. A model for the interaction between PI31 and Fbxo7 is presented.

Acknowledgements

The work presented in this thesis would not have been possible without the help and support of many people. Firstly I would like to thank my supervisor, Neil McDonald, for allowing this Physics student the opportunity to learn crystallography and for his guidance and support throughout the years, it has been invaluable. Without the help of members of the Structural Biology Lab both past and present this thesis would have been a very different endeavour, thanks for keeping me sane in the lab, during the long nights at the synchrotron as well as for joining me for drinks in the George. I would especially like to thank Phillip Knowles for all his work on this project as well as his almost encyclopaedic knowledge of ways to produce useful protein and constructs. Without the help of Judith Murray-Rust and Matthew Newman I may never have emerged from the maze of data collection and analysis to produce a fully refined structure, thank you both so much for your time, knowledge and patience.

Thanks to Jane Kirk, Jon Moore, Tod Duncan and Heike Laman for their help with the removal of any "becky-isms" from this thesis and for all their helpful comments and suggestions.

This work was supported by the Barbara Mary Hill fellowship and I would like to thank the trustees of this fund for supporting the work of Cancer Research UK and my work in particular. I hope that you will be pleased with the results.

I have had a great four years in London thanks mostly to my friends. Thanks guys for all the nights out and the chats, especial thanks to Emily-Jane for being the best friend a girl could ask for.

The love and support of my family has been a constant in my life and without it I wouldn't have been able to do this, thank you Mum, Dad and Katy.

Finally, thanks Dave for your love, support and understanding; this thesis is dedicated to you.

Declaration

The following figures were kindly provided by Dr. Heike Laman, Wolfson Institute, UCL, London.

Figure 2.11 – Fbxo7 forms a complex with Skp1, Cul1 and Rbx1

Figure 2.12 – Identification of PI31 as a binding partner of Fbxo7.

Figure 5.5 – Identification and testing of possible lysine residues for ubiquitination

Figure 5.6 – PI31–Fbxo7-Skp1 interaction is required for PI31 ubiquitination but does not target PI31 to the proteasome

Figure 5.7 – PI31 co-localises with the ER

Figure 5.8 – Expression of PI31 and Fbxo7 in various tissues and cell lines

Phillip Knowles, Structural Biology Laboratory, Cancer Research UK, provided the following data.

Figure 2.3 – Limited proteolysis of Fbxo7(129-398)-Skp1(1-146)

Figure 2.6 – Constructs produced and expression trials for Fbxo7-Skp1 complexes (SDS-PAGE of F1 construct only)

Figure 2.9 – Crystals of $Fbxo7^{F3}$-Skp1

A list of all donated constructs can be found in section A.3.1.

Table of contents

List of figures

List of tables

Abbreviations

2-ME	2-mercaptoethanol
3D-PSSM	3 dimensional position specific scoring matrix
AMP	Adenosine monophosphate
APC	Anaphase-promoting complex
APS	Ammonium persulphate
ATP	Adenosine triphosphate
AU	Asymmetric unit
AUC	Analytical ultracentrifugation
BTB	Broad complex, Tramtrack, Bric-a-brack
CCD	Charge-coupled device
CCP4	Collaborative Computational Project Number 4
CD	Circular dichroism
cDNA	Complementary DNA
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
$DiOC_6(3)$	3, 3'-Dihexyloxacarbocyanine, iodide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
dTTP	2'-deoxythymidine 5'-triphosphate
EBI	European Bioinformatics Institute
EMTS	Ethylmercurithiosalicylic acid
ER	Endoplasmic reticulum
ESRF	European Synchrotron Radiation Facility

EtBr	Ethidium Bromide
FPLC	Fast protein liquid chromatography
GST	Glutathione-S-transferase
HECT	Homologous to E6-AP Carboxy Terminus
HEPES	N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
HMM	Hidden Markov modelling
HURP	Hepatoma up-regulated protein
HVS	*Herpesvirus samurai*
IFN-γ	Interferon-γ
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
Kb	Kilobase
KDa	KiloDaltons
L	Litre
LRR	Leucine rich repeat
MAD	Multiple anomalous diffraction
mAU	Milliabsorbance unit
MCS	Multiple cloning site
MES	2-Morpholinoethanesulfonic acid
mg	Milligram
MHC	Major histocompatibility complex
ml	Millilitre
mRNA	Messenger RNA
NCS	Non crystallographic symmetry
NEDD8	Neuronal-precursor-cell-expressed developmentally downregulated protein-8

NMR	Nuclear magnetic resonance
NTD	N-terminal domain
Ntn	N-terminal nucleophile
ori	Origin of replication
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
PGPH	Peptidylglutamyl-peptide hydrolytic
pI	Isoelectric point
PMSF	Phenylmethylsulphonylfluoride
POZ	Poxvirus and zinc finger
PP_i	Inorganic pyrophosphatase
PRR	Proline rich region
RBS	Ribosome binding site
RING	Really Interesting New Gene
RMSD	Root mean square deviation
RNA	Ribonucleic acid
rpm	Revolutions per minute
Rpn	Regulatory particle non-ATPase
Rpt	Regulatory particle ATPase
rRNA	Ribosomal RNA
SAD	Single anomalous diffraction
SCF	Skp-Cullin-Fbox
sd	Standard deviation
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium-dodecyl sulphate-polyacrylamide gel electrophoresis

Se-Met	Selenomethionine
SIR	Single isomorphous replacement
SSM	Secondary structure matching
SUMO	Small ubiquitin-like modifier
TCR	T cell receptor
TEV	Tobacco etch virus
Tris	Tris[hydroxymethyl]aminomethane
tRNA	Transfer RNA
Ubl	Ubiquitin like domain
UbLPs	Ubiquitin like proteins
UIM	Ubiquitin interacting motif
UV	Ultra-violet
µl	Microlitre
µM	Micromolar

1 Introduction

1.1 Protein degradation through the ubiquitin-proteasome pathway

Since the discovery in 1979 that a specific protein marker, called ubiquitin, tags and targets proteins for degradation (Ciechanover *et al.* 1980; Hershko *et al.* 1980) it has become clear that the degradation of cellular proteins is a highly complex, temporally controlled, and tightly regulated process. The ubiquitin-proteasome system plays major roles in a variety of basic pathways during cell life and death through cellular processes such as transcription regulation (Muratani and Tansey 2003) and antigen presentation (Shastri *et al.* 2002). In 2004 the Nobel Prize in Chemistry was awarded to Aaron Ciechanover, Avram Hershko and Irwin Rose in recognition of their initial discovery of this process.

Protein ubiquitination is a tightly controlled process, which is required to target proteins for destruction by a multi-subunit ATP-dependent protease called the proteasome. The proteasome consists of a central catalytic machine, known as the 20S proteasome and two terminal regulatory sub-complexes that are attached to both ends of the central portion to form the active proteasome. Structure determination of the 20S proteasome and its regulators has revealed much of its function (for a review of this see (Bochtler *et al.* 1999)). Regulation of the proteasome itself is achieved by restricting the proteolytic activity to specific locations that can be accessed only by those polypeptides destined for destruction. The best characterised chambered protease is the proteasome but bacteria, archaea and eukaryotes all use the simple method of restrictive access for regulation of proteases (Groll and Clausen 2003).

Synthetic inhibitors of the proteasome, such as the drug Bortezomib, have been shown to have *in vivo* anti-tumour efficacy (Voorhees *et al.* 2003) and have recently been given approval for use in the treatment of multiple myeloma. It appears that transformed cells are more susceptible to proteasome inhibition than normal cells. A key factor in the ability to kill certain cancer cells appears to be the ability to block activation of NF-κB, which is constitutively active in myeloma cells. NF-κB has anti-apoptotic effects and so treatment with proteasome inhibitors can sensitise cells to other anti-cancer treatments (Goldberg and Rock 2002).

The ubiquitin-proteasome system is involved in many cellular processes and one well characterised example is the cell cycle (Figure 1.1). For the cell cycle to proceed correctly specific proteins must be produced at critical times, these proteins must be localised correctly (Moore *et al.* 2003) and then degraded at specific points in the cell cycle in a ubiquitin-dependent manner (Evans *et al.* 1983; Glotzer *et al.* 1991). The periodic but complete destruction of signalling proteins during this process ensures that the cycle continues in its progression in a timely manner without returning to previous phases.

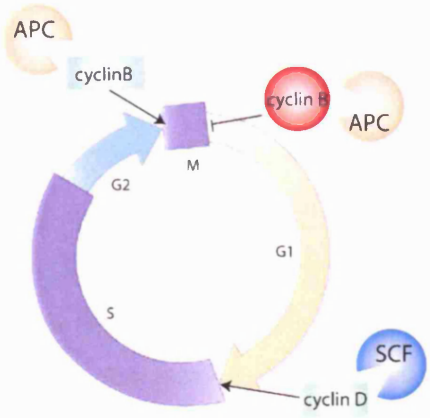

Figure 1.1 - Cyclins targeted by ubiquitin-mediated proteolysis.

The G1 (gap phase 1), S (DNA synthesis), G2 (gap phase 2) and M (mitosis) phases of the mammalian cell cycle are depicted. Positive cell-cycle regulators are shown as green rectangles, whereas negative regulators are depicted as red circles.

Proteins in sea urchins and clam embryos, called cyclins, were found to be continuously synthesized and accumulated throughout the cell cycle but abruptly disappeared during mitosis (Evans *et al.* 1983). The striking behaviour of these proteins indicated that cyclins might be involved in the regulation of cell division. The importance of this discovery was recognised in 2001 by the joint award of the Nobel Prize in physiology and medicine to Tim Hunt. The rapid disappearance of one of these cyclins, cyclin B, was later revealed to be dependent on the ubiquitin system (Glotzer *et al.* 1991). Though cyclin B is a positive regulator of the G2-M phase transition of the cell cycle the degradation of this protein is required for exit from mitosis (Glotzer *et al.* 1991). The rapid exit from mitosis that is afforded by the irreversible destruction of cyclin B is crucial for cellular and genomic integrity (Wheatley *et al.* 1997; Jin *et al.* 1998).

The D type cyclins are also regulated by the ubiquitin system. D type cyclins are rate-limiting controllers of G1 phase progression. The elimination of the intrinsically instable D cyclins is initiated by phosphorylation on a specific site (reviewed in (Lee and Yang 2003) and (Sherr 1995)). In addition to this it has been shown that cyclins A and E are also substrates of the ubiquitin-proteasome system. Despite the apparent complexity of the cell cycle the underlying control is through phosphorylation and phosphorylation-dependent degradation through the ubiquitin-proteasome system.

1.2 Ubiquitin and ubiquitin-like proteins

Proteins are targeted to the proteasome for destruction by covalent ligation to ubiquitin, a 76 amino acid protein (Hershko *et al.* 1980). The role of ubiquitin in protein degradation was discovered and the main enzymatic reactions of the process elucidated by biochemical studies using a cell free system ((Ciechanover *et al.* 1980) and reviewed in (Hershko and Ciechanover 1998)). The ligation of ubiquitin to a protein requires the sequential action of three enzymes (E1, E2 and E3). These enzymes were originally identified by affinity chromatography of a crude reticulocyte extract on ubiquitin-Sepharose by Hershko *et al* in 1983 (Hershko *et al.* 1983). The C-terminal glycine residue of ubiquitin is activated by an activating enzyme (E1) in a step that requires ATP (Step 1, Figure 1.2). This step involves the intermediate formation of a ubiquitin adenylate in which the carboxyl-terminal glycine of ubiquitin and AMP are in an acyl-phosphate linkage (Haas and Rose 1982; Haas *et al.* 1983) along with release of PP_i and the binding of ubiquitin to a cysteine residue of the E1 enzyme by a thiolester linkage and the release of AMP. This activated ubiquitin is then transferred to the active site cysteine residue of a ubiquitin conjugating enzyme (E2) (Step 2, Figure 1.2). In the final step, which is catalysed by a ubiquitin-protein ligase (E3), ubiquitin is linked by its C-terminus in an amide isopeptide linkage to a ε-amino group of the substrate protein's lysine residues (Step 3, Figure 1.2). In rare cases ubiquitin can be conjugated to the N-terminal amino group of the substrate (reviewed in (Ciechanover and Ben-Saadon 2004)). By successively adding activated ubiquitin moieties to lysine residues on the previously conjugated ubiquitin molecule, a polyubiquitin chain is synthesised. The three steps outlined in Figure 1.2 are the simple, progressive mechanism used in the ubiquitination cascade independently of whether the substrate bound ubiquitin(s) signal proteasomal proteolysis or directs the protein to another fate.

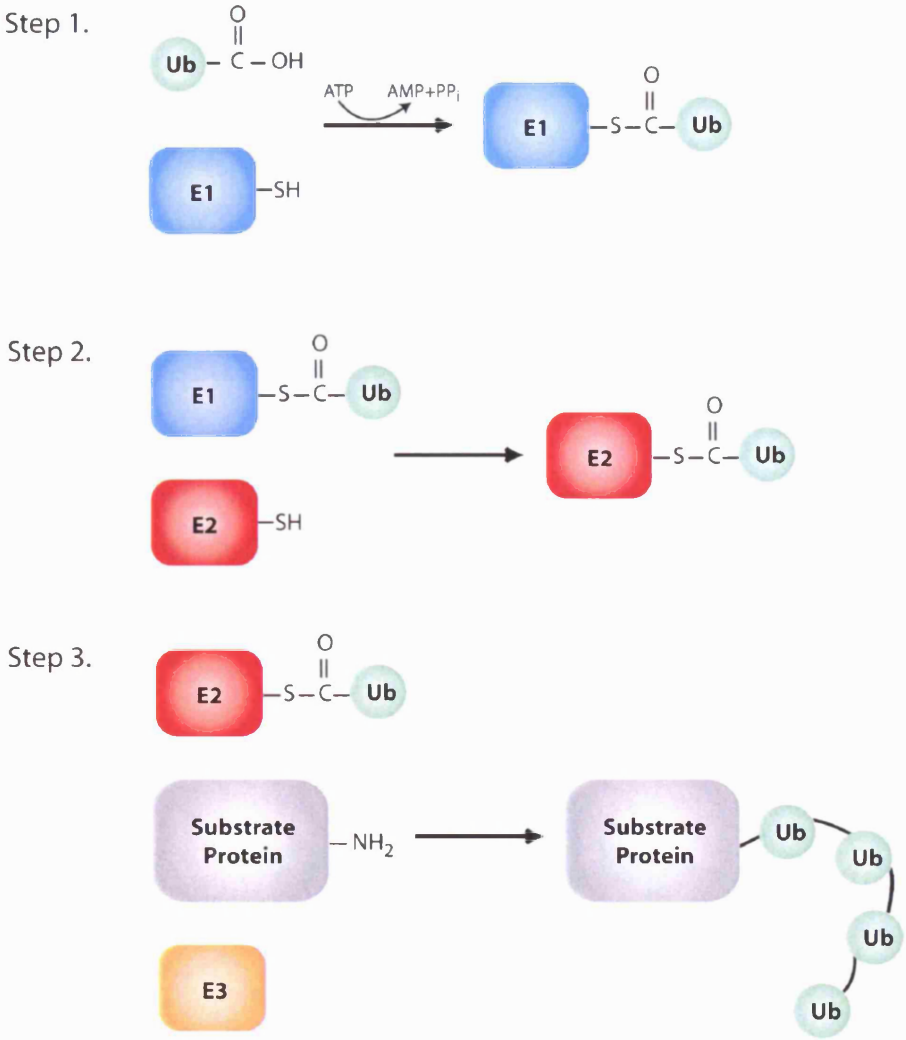

Figure 1.2 – Enzymatic reactions of the ubiquitin system.

Step 1. E1 enzyme activates ubiquitin in an ATP-dependent reaction. Step 2. E1 enzyme transfers the activated ubiquitin to the E2 ubiquitin conjugating enzyme. Step 3. E2 transfers the ubiquitin to the substrate protein in cooperation with the E3 ubiquitin ligase. SH refers to the free form of catalytic cysteine and PP_i to inorganic pyrophosphate.

It has been recently proposed that in addition to the well established E1-E2-E3 cascade a supplementary enzyme may be required in some cases. This enzyme would allow the elongation of ubiquitin chains on monoubiquitinated substrates to achieve the polyubiquitin chains required for recognition by the proteasome. These enzymes have been termed E4 enzymes and are reviewed in (Hoppe 2005).

Ubiquitin has a number of related homologues known collectively as ubiquitin-like proteins (referred to here as UbLPs) some of which also use the basic set of catalytic reactions that are involved in ubiquitination (outlined in Figure 1.2). Not all UbLPs require an E3-like ligase to assist the binding of the UbLP to its target substrate for example the E2 enzyme for SUMO-1 can directly both bind and sumoylate its substrate *in vitro* although whether this occurs *in vivo* is not clear (Schwartz and Hochstrasser 2003).

By tagging proteins with ubiquitin or ubiquitin-like proteins the cell can create a diverse family of post translationally modified proteins that can be identified by downstream protein receptors and used to control many regulatory pathways in the cell (Glickman and Ciechanover 2002).

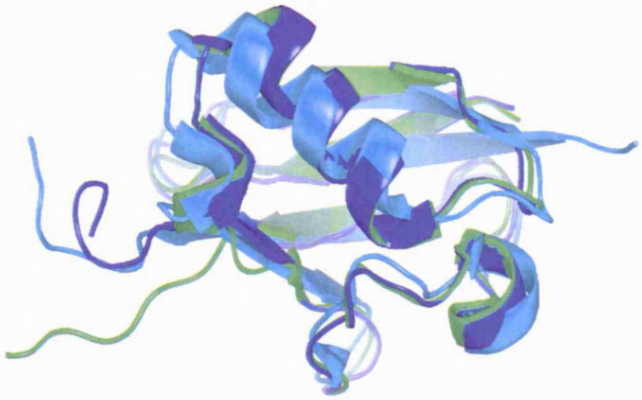

Figure 1.3 – The ubiquitin superfold

An overlay of ubiquitin (blue; PDB accession code 1OGW (Vijay-Kumar *et al.* 1987)), SUMO-1 (cyan; PDB accession code 1WYW (Baba *et al.* 2005)) and NEDD8 (green; PDB accession code 1R4M (Walden *et al.* 2003)), produced using PyMol (DeLano 2004).

Ubiquitin and ubiquitin-like proteins all contain the ubiquitin superfold (Figure 1.3), which is a β-grasp fold, but sequence homology between the proteins varies considerably. In the course of evolution UbLPs have evolved to give the present

repertoire of small proteins with distinct functions, which are only now starting to be understood.

Substrate targeting to the proteasome requires the assembly of a polyubiquitin chain formed by an isopeptide bond between Lys48 of one ubiquitin moiety and the C-terminal carboxyl group of the next ubiquitin in the chain. By contrast, ubiquitin chains with non-Lys48 linkages or even a single ubiquitin can sometimes be attached to a substrate and may direct its target to a different fate. The most common alternative to Lys48 is Lys63 and Lys63-linked polyubiquitination has been implicated in many cellular events that do not require destruction via the proteasome (reviewed in (Hochstrasser 2004)).

Lys48-linked polyubiquitin chains themselves do not always target the substrate protein for destruction (Hochstrasser 2004). The transcription factor Met4 is a stable protein even after Lys48-linked polyubiquitination and indeed transcriptional repression of Met4 target genes is dependent on the ubiquitination of Met4 (Kaiser *et al.* 2000; Flick *et al.* 2004). This reveals a proteolysis-independent function for Lys48-linked polyubiquitin chains.

Modification of substrate proteins by other UbLPs also has a wide variety of functions within the cell by changing the stability, localisation or activity of the target protein. Most of these functions are only starting to be established, with SUMO-1 and NEDD8 being perhaps the best understood UbLPs at this time (Aguilar and Wendland 2003).

In addition to ubiquitin and UbLPs, which are small single subunit proteins, ubiquitin-like (Ubl) domains can also occur as stable elements within multi-domain proteins, although these Ubl domains cannot be processed or conjugated. Ubl domains, which usually have an N-terminal location, have significant sequence similarity to ubiquitin and display its characteristic fold. These Ubls can often function as targeting elements and have been shown to bind to the proteasome (Walters *et al.* 2002).

1.3 E3 ubiquitin ligases

There is a hierarchical structure within the ubiquitin cascade: there is one E1 enzyme (McGrath *et al.* 1991); a limited number of E2 enzymes each of which may interact with several E3 enzymes (Pickart and Rose 1985); and a much larger number of E3 enzymes.

The level of complexity of control of this system therefore increases throughout the pathway.

The large number of E3 ubiquitin ligases confer the specificity, regulation and complexity on this pathway by mediating the transfer of ubiquitin from the E2 ubiquitin conjugating enzyme to the substrate protein. The ability of E3 proteins to recognise and catalyse the processive polyubiquitination of their substrate protein(s) is vital for the protein to be targeted to the proteasome because if the correct number of ubiquitin moieties is not transferred to the substrate it will not be recognised by the proteasome for degradation.

Almost all known E3 ligases contain either a HECT domain or a RING domain (Figure 1.4) and are classified accordingly. The major difference between these ligases is that the HECT domain E3 ubiquitin ligases form a covalent bond with ubiquitin through an active site cysteine (Huibregtse *et al.* 1995) whereas RING domain E3 ubiquitin ligases do not. A further sub-classification within the RING E3 ubiquitin ligases can be made between single subunit RING E3 ubiquitin ligases, the multi-subunit SCF-like RING E3 ubiquitin ligases and the multi-subunit APC RING E3 ubiquitin ligases.

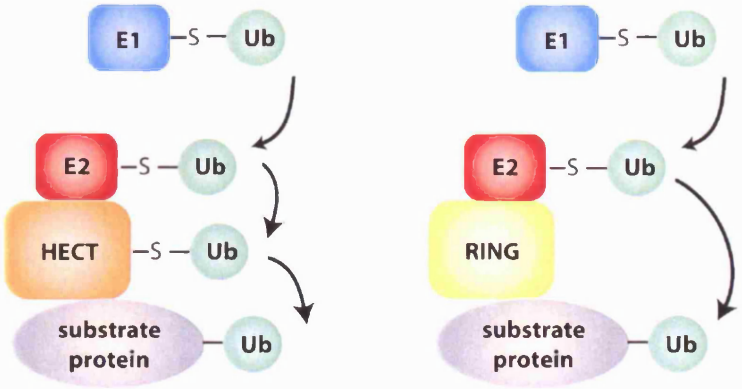

Figure 1.4 – Two families of ubiquitin protein ligases

HECT domain E3 ubiquitin ligases form a covalent bond with ubiquitin through an active site cysteine whereas RING domain E3 ligases mediate the transfer of ubiquitin to the substrate through a different mechanism.

The focus of this introduction is on the SCF subgroup of the RING E3 ubiquitin ligases. Excellent reviews have been published elsewhere on the HECT E3 ligases (Pickart 2001), the single subunit RING E3 ubiquitin ligases (Jackson *et al.* 2000) and the multi-subunit APC RING E3 ubiquitin ligases (Vodermaier 2004; Castro *et al.* 2005).

1.3.1 SCF-like E3 ubiquitin ligases

All non-HECT domain E3 ligases promote the transfer of ubiquitin from the E2 without forming a covalent interaction with ubiquitin and they all contain either a RING finger domain or the structurally equivalent U-box domain (Hatakeyama and Nakayama 2003). The RING (really interesting new gene) finger domain displays a series of histidine and cysteine residues with a characteristic spacing (Lovering *et al.* 1993) that allows for the co-ordination of two zinc ions in a cross-brace structure. Whereas spacing of the two ions is conserved throughout the family, which consists of many hundreds of proteins, the primary sequence is not conserved. RING finger proteins act as molecular scaffolds that bring other proteins together (Pickart 2001). The RING finger domain is thought to bring the E2 enzyme into contact with the RING domain E3 ligases (Zheng *et al.* 2000).

The largest subgroup of the RING finger containing E3 ligases is the SCF-like E3 ligases (reviewed in (Deshaies 1999)). The SCF E3 ubiquitin ligases consist of a complex of the Skp1-Cullin1-F-box proteins bound to the E2 binding RING finger protein Rbx1 (Kamura *et al.* 1999) (Figure 1.5).

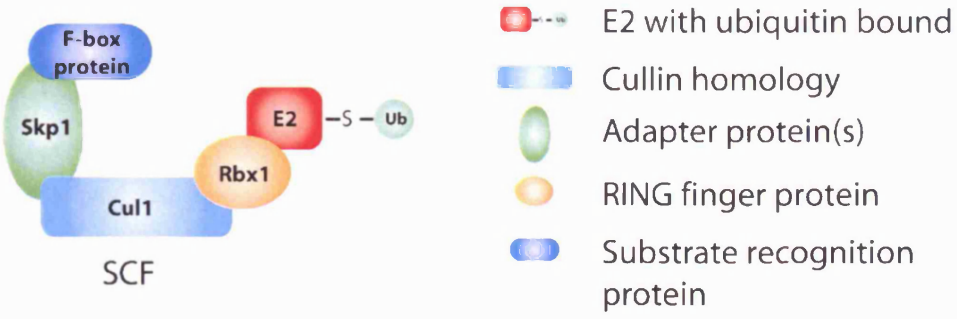

Figure 1.5 – The SCF RING E3 ligase

The interaction between Skp1 and F-box proteins was first observed between Skp1 and cyclin F. A 40 amino acid motif was identified in cyclin F that mediates the interaction with Skp1 by Bai and co-workers (Bai *et al.* 1996). This 40 amino acid motif is known as the F-box domain and it is present in a large family of multi-domain F-box proteins that also contain protein-protein interaction domains. The most common protein-protein interaction domains in F-box proteins are WD40 domains and leucine rich repeats (LRR) and so the nomenclature for newly discovered F-box proteins is Fbxw (for WD40 repeat containing proteins), Fbxl (for LRR containing proteins) and Fbxo (for all other F-box proteins) (Cenciarelli *et al.* 1999; Winston *et al.* 1999). The large

numbers of F-box proteins in all the genomes predicted that these proteins conferred specificity on the SCF ligase and add to the complexity of the pathway. Skowyra *et al* (Skowyra *et al.* 1997) showed that this is true in yeast for the F-box protein cdc4 and its substrate protein phosphorylated Sic1. Each F-box protein binds to several substrates and so many hundreds of proteins can be targeted to the proteasome by interaction with the SCF E3 ubiquitin ligase through the variable F-box domain. The specific complexes were designated SCF for Skp1-Cullin1-F-box protein, with the F-box protein indicated by a superscript, as in SCF^{cdc4} or $SCF^{cyclin\ F}$.

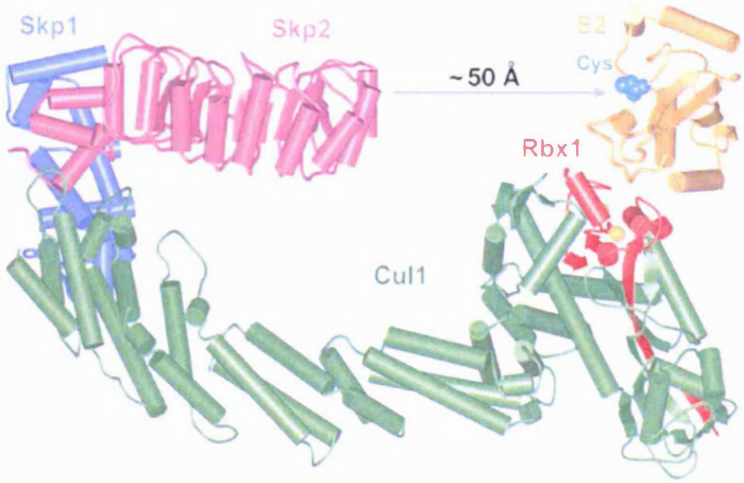

Figure 1.6 – Model of the SCF^{Skp2}-E2 complex

Cul1 (green), Rbx1 (red), Skp1 (blue), Skp2 (pink), E2 (yellow) are shown in cartoon representation. The zinc atoms in the Rbx1 are shown as yellow spheres and the active site cysteine of the E2 enzyme where the ubiquitin would be attached is shown in space filling representation and coloured in cyan. The grey arrow indicates the 50Å gap between the F-box protein and the active site cysteine where the substrate is thought to bind. Figure reproduced from (Zheng *et al.* 2002).

The structure solution of the Cul1-Rbx1-Skp1-F-box^{Skp2} revealed the organisation of the SCF ligase and provided valuable insights into the mechanism of ubiquitin transfer (Zheng *et al.* 2002). Cullin1 is an elongated protein that consists of a long stalk made of three repeats of a five helix bundle called the cullin-repeat motif and a globular domain that binds the RING finger protein Rbx1 through an intermolecular β-sheet. The Skp1-F-box substrate recognition complex binds to the tip of the Cullin1 stalk. Cullin1 is a rigid scaffold that organises the Rbx1, Skp1 and F-box protein subunits, positioning the substrate for optimal ubiquitin transfer. This crystal structure was used to produce a model using the previously solved Skp1-Skp2 complex (Schulman *et al.* 2000) and the

RING finger – E2 complex (Zheng *et al.* 2000) to reveal the overall structure of the E2-SCF^{Skp2} interaction (Figure 1.6). The model revealed a 50Å gap between the F-box protein and the active site cysteine of the E2 enzyme to allow the binding of the substrate and presentation to the E2 ligase for the transfer of ubiquitin to the substrate protein. This proposed mechanism was further revealed by two crystal structures of Skp1-F-box protein complexes which contained peptides of the substrates (SCF^{cdc4} (Orlicky *et al.* 2003) and $SCF^{\beta\text{-}TrCP}$ (Wu *et al.* 2003)) which both show that a peptide of the substrate protein is bound to the F-box protein in a position to allow access to the active site cysteine of the E2.

The SCF is not constitutively active and it is controlled by NEDD8 modification (neddylation) of Cullin1 (Read *et al.* 2000). Cand1 protein binds to the unmodified Cullin1-Rbx1 complex and occludes Skp1, thereby preventing the assembly of an active SCF complex. Neddylation of Cullin1 prevents the binding of Cand1 and stabilises the active Skp1-Cul1-Rbx1 complex. Cand1 can dissociate Skp1 and NEDD8 can dissociate Cand1 and so the inactive and active complexes exist in equilibrium (Zheng *et al.* 2002).

The Cullin-RING-adaptor theme is repeated in at least five other multi-subunit ubiquitin ligases. The complicated combinations of proteins belie a coherent nomenclature and so here SCF-like ubiquitin ligases will be named with the original nomenclature for the SCF. This is based on the order of the Skp1-like linker, Cullin and the adaptor class for example, an EloC-EloB-Cul2-Rbx1-SOCS box protein complex is called EC_2S. An outline of all the SCF-like protein complexes is given in Table 1.1 and each of these will be discussed briefly, for a more complete review see (Willems *et al.* 2004).

	SCF	**EC_2S**	**BC_3B**	**SC_4F?**	**SC_7F?**
Cullin domain protein	Cul1	Cul2 or Cul5	Cul3	Cul4	Cul7
Adaptor proteins	Skp1	Elongin C, Elongin B	BTB/POZ domain proteins	?	Skp1
Substrate binding proteins	F-box proteins	SOCS box proteins	BTB/POZ domain proteins	?	F-box proteins (Fbxw6, Fbxo29)
Activator proteins	NEDD8	NEDD8	NEDD8	NEDD8	NEDD8

Table 1.1 – Components of SCF-like ubiquitin E3 ligases

Information taken from (Passmore and Barford 2004) with BTB meaning Broad complex, Tramtrack, Brick-a-brack and POZ Poxvirus and Zinc finger.

A class of SCF-like complexes that has been shown to be closely related to the SCF is built on the Elongin C (EloC)-Cul2 core complex. The EloC forms a heterodimer complex with the ubiquitin-like protein Elongin B (EloB) and these two proteins act as an adaptor between the substrate binding proteins and the Cul2-Rbx1 complex. The substrate binding proteins contain a 40 amino acid motif called the SOCS box as well as further protein-protein interaction domains which is reminiscent of the F-box proteins. The crystal structures of the SOCS box protein VHL bound to EloB/C (Stebbins *et al.* 1999) and the F-box protein Skp2 bound to Skp1 (Schulman *et al.* 2000) reveal striking similarities (Figure 1.7). The SOCS box and F-box both consist of a three helix bundle, the N-terminus of Skp1 resembles the structure of Elongin C. The overall layout of the adaptor proteins is the same and therefore substrate binding to the protein is likely to be similar (Kile *et al.* 2002).

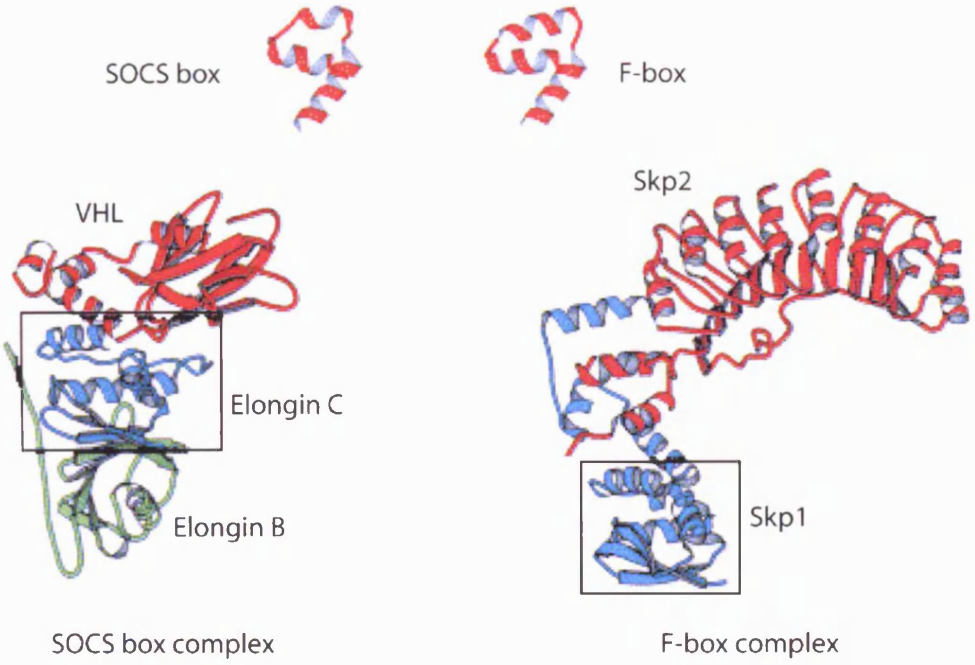

Figure 1.7 – Structural comparison of EC_2S and SCF complexes

Elongin C and Skp1 are shown in blue, Elongin B is shown in green and the SOCS and F-box proteins are shown in red. The boxed area highlights the structural similarity between the N-terminus of Skp1 and Elongin C. Figure adapted from (Kile *et al.* 2002).

A third type of SCF-like E3 ligases consist of BTB-containing proteins acting as both a Skp1-like adaptor and F-box protein-like substrate specificity factors in association with Cul3 and Rbx1 (Pintard *et al.* 2003; Xu *et al.* 2003). These complexes are termed BC_3B for BTB-Cul3-BTB reflecting the dual role of the BTB protein. A structural

comparison of the Skp1 and EloC structures revealed that there is a BTB fold embedded within them (Stebbins *et al.* 1999; Schulman *et al.* 2000) emphasising the structural link between all these SCF-like E3 ubiquitin ligases. BTB proteins can also contain protein-protein interaction domains, which could interact with substrate proteins in a way homologous to F-box proteins (Stogios and Privé 2004).

The other possible SCF-like E3 ligases are less well characterised with, for example, several SOCS box proteins (including VHL) binding Cul5 as well as Cul2. Cul4A may also act in E3 ligases and Cul7 forms an SCF-like E3 ligase with Rbx1, Skp1 and the F-box protein Fbxw6/Fbxo29. This interaction between the subunits of the different RING E3 ligases reveals the possibility for cross-talk between these ligases. These interactions add an additional layer of complexity to this system, which is required because aberrant degradation of proteins might have serious consequences.

1.3.2 Signals that target proteins for ubiquitination

The E3 ubiquitin-protein ligases use different substrate specificity factors to bind to their substrates. These differences allow specific proteins to be conjugated with ubiquitin at the required time and place. The substrate recognition proteins identify signals within the substrate that include post-translational modifications and motifs within the protein itself. In contrast to the complexity of the enzymes involved in the recognition of these substrates the protein signals themselves are very simple.

Of the SCF-like E3 ligases the best understood is the SCF itself and some of the signals recognised by this ligase are well characterised. The F-box proteins recognise at least two different forms of post-translational modification in substrate proteins. F-box proteins were originally thought to bind only phosphorylated substrates through their C-terminal protein-protein interaction domains (Skowyra *et al.* 1997). This is an elegant and simple solution as SCF complexes can be thought of as coupling protein kinase signalling to control the abundance of active proteins. More recently it has been reported that some of the Fbxo type F-box proteins interact with their substrates by recognising sugar chains on the substrate protein with the substrates of these F-box proteins being recognised in an N-glycosylation-dependent manner (Yoshida *et al.* 2002; Mizushima *et al.* 2004). The discovery of this mechanism for one class of Fbxo type F-box proteins indicates that there are likely to be other signals recognised by F-box proteins that are, as yet, uncharacterised.

Studies of the other classes of SCF-like ligases implicate phosphorylation as a targeting signal. In the EC_2S protein ligase the substrate proteins are only recognised following post-translational modification by phosphorylation (Kile *et al.* 2002) and in the BC_3B protein ligase it is also possible that phosphorylation is required for substrate recognition as in the case of the BTB protein Mel26 and its substrate Mei1 (Willems *et al.* 2004).

1.3.3 F-box protein, Fbxo7

Fbxo7 is an F-box protein of the Fbxo7 class which Cenciarelli and co-workers (Cenciarelli *et al.* 1999) first identified in a yeast two-hybrid screen using human Skp1 as bait. Biomolecular analysis revealed that Fbxo7 contained the F-box domain and in addition to this it also contained proline rich regions (PRR) and a leucine zipper, which are both potential protein-protein interaction domains. They went on to show that FLAG-tagged Fbxo7 is a component of an SCF complex *in vivo* by immunoprecipitation with anti-FLAG antibody. This interaction is dependent on the presence of a functional F-box domain in Fbxo7. SCF^{Fbxo7} was also shown to have ubiquitin ligase activity in the presence of an E1 and E2 enzyme. This was the first demonstration that F-box proteins from the Fbxo family could associate with ubiquitin ligase activity.

Independently of the work done by Cenciarelli *et al*, another group identified Fbxo7 to be a predicted Fbxo type F-box protein with 522 amino acids and an F-box motif located in the C-terminal half of the protein (Ilyin *et al.* 2000). Fbxo7 mRNA was reported to be ubiquitously expressed but enriched in bone marrow, foetal liver, testis and thyroid glands. This led Ilyin *et al* (Ilyin *et al.* 2000) to suggest that, as the highest expression is in the main hematopoietic tissues, this could imply a role for Fbxo7 in the regulation of hematopoiesis.

The work by Cenciarelli *et al*, which showed that Fbxo7 is part of a functional E3 ligase, did not point to any substrate proteins for this complex. The first reported substrate for SCF^{Fbxo7} was the Cdk1-cyclin B-phosphorylated hepatoma up-regulated protein (HURP) (Hsu *et al.* 2004). Hsu *et al* (Hsu *et al.* 2004) demonstrated that Fbxo7 functions as an adaptor linking the core SCF complex with its substrate, HURP, through the PRR of Fbxo7. They were also able to show that this interaction is dependent on the phosphorylation of HURP on multiple sites by Cdk1-cyclin B. The kinase Aurora A

has also been shown to phosphorylate HURP (Yu *et al.* 2005) and so may also be able to target HURP for binding to SCF^{Fbxo7}. HURP is a cell cycle regulated gene that is over expressed in hepatocellular carcinoma (Tsou *et al.* 2003). The endogenous levels of HURP mRNA are tightly regulated during the cell cycle and its expression is elevated in the G2-M phase. Cells stably transfected with HURP appear to have the characteristics of tumour cells with a reduced dependence on extra cellular growth factors and the potential to exhibit anchorage-independent growth. Tsou *et al* (Tsou *et al.* 2003) suggested that over expression of HURP in hepatocellular carcinoma may be a consequence of deregulation which can lead to cancer by allowing cells to progress through the cell cycle. This protein has also been associated with colon cancer, breast cancer (Bassal *et al.* 2001) and transitional cell carcinoma (Chiu *et al.* 2002). This uncontrolled expression of HURP means that the identification of the SCF ligase responsible for the timed degradation of the protein could reveal a possible way to revert these cancer cells to a more normal phenotype. These results are further encouraging as Ilyin *et al* (Ilyin *et al.* 2000) reported Fbxo7 expression in the liver which could imply that one possible *in vivo* role of SCF^{Fbxo7} could be the control of HURP protein levels.

Fbxo7 was also reported to be a putative oncogene by our collaborators Laman *et al* (Laman *et al.* 2005). This work identified Fbxo7 as interacting with viral D-type cyclins in a yeast two-hybrid screen of a human cDNA library using a viral D-type cyclin as bait. Fbxo7 interacts with both cellular and viral D cyclins and specifically binds Cdk6, *in vitro* and *in vivo*. The interaction with Fbxo7 is specific to D cyclin/Cdk6 complexes, as Fbxo7 does not interact with either cyclin E or Cdk4. An Fbxo7 knockdown decreased Cdk6 association with cyclin D and over expression of Fbxo7 increased D cyclin/Cdk6 activity. Over expression of Fbxo7 transformed murine fibroblasts and rendered them tumorigenic in athymic nude mice. These transformed phenotypes were dependent on Cdk6 activity, as concomitant knockdown of Cdk6 eliminated the changes. Fbxo7 was shown to be highly expressed in epithelial tumours, but not in normal tissues, suggesting it may have a proto-oncogenic role in human cancers (Laman *et al.* 2005).

The connection between F-box proteins and cancer has been well established through the study of other F-box proteins such as Skp2. Skp2 has been shown to be responsible for the activation and stabilisation of an oncogene, c-Myc (Kim *et al.* 2003; von der

Lehr *et al.* 2003), as well as playing a role in the degradation of the tumour suppressors p27 (Carrano *et al.* 1999; Tsvetkov *et al.* 1999), p21 (Bornstein *et al.* 2003), p57 (Kamura *et al.* 2003) and p130 (Tedasco *et al.* 2002). These latter interactions indicate that the Skp2 protein itself may be the product of a proto-oncogene. Several studies have revealed that Skp2 expression significantly and directly correlates with tumour malignancy and aggressiveness and is associated with poor prognosis in human lymphomas, prostate cancers and ovarian adenocarcinomas ((Guardavaccaro and Pagano 2004) and references therein).

It has been suggested as a consequence of the strong link between the SCF complexes and cancer that individual F-box proteins such as Skp2 may be a target for treatment as well as being an effective prognostic marker (Guardavaccaro and Pagano 2004). The diversity of pathways regulated by the ubiquitination system means that targeting the proteins higher in the chain would be likely to disrupt too many essential processes and so the substrate specificity factors, the F-box proteins, would be ideal drug targets for new cancer therapies. As discussed previously Fbxo7 has been shown in two different reports to be implicated in the control of cell cycle proteins, HURP and cyclin D/Cdk6, and in both reports Fbxo7 displayed proto-oncogenic behaviour. It is possible that Fbxo7 represents one such suitable cancer drug target. Further work would be required to assess the suitability of Fbxo7 as a drug target as protein-protein interactions are complex targets. The structure determination of Fbxo7 would greatly assist this analysis as well as facilitating any possible drug design.

1.4 Structure and function of the proteasome

1.4.1 Architecture of the proteasome

After proteins are modified by ubiquitin they can be targeted for destruction by the proteasome. The proteasome is found throughout all three kingdoms, bacteria, archaea and eukaryotes, and has a similar core structure. The archaeal proteasome complex consists of two different subunits, α and β, which form a barrel-shaped structure of four stacked rings (Löwe *et al.* 1995) (Figure 1.8). The two inner rings consist of seven β subunits each and the two outer rings consist of seven α subunits each. The overall architecture of the proteasome reveals an elongated cylinder with a central penetrating channel with large cavities and narrow constrictions.

Structural studies on the *Thermoplasma acidophilum* proteasome (Löwe *et al.* 1995) have shown that the α and β subunits are similar, consisting of a core of two five-stranded anti-parallel β-sheets that is flanked by three α helices on the top and two on the bottom (Figure 1.8). Sequence analysis of the α and β subunits indicates that they evolved from a common ancestral gene with the main difference being the N-terminal extension of ~35 amino acids (Löwe *et al.* 1995). These amino acids form the additional N-terminal α helix seen in the α subunits and are not present in the β subunits. Instead of this N-terminal α helix the β subunits contain a prosequence of various lengths, which are removed during proteasome assembly. These prosequences aid the efficient incorporation of the β subunits into the proteasome by a co-operative mechanism (Schmidt *et al.* 1999).

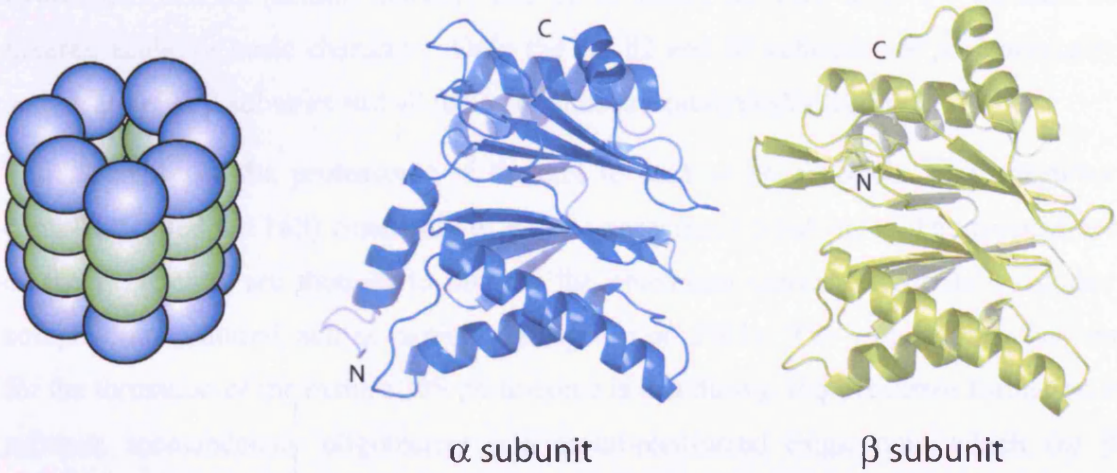

Figure 1.8 – Representation of the 20S proteasome and the α and β subunits

The 20S proteasome consists of four stacked rings of alpha (blue) and beta (green) subunits. The 20S proteasome of *T. acidophilum* is shown with each subunit as a sphere. The cartoon representations of the alpha and beta subunits show that they are similar in structure with the main difference being the N-terminal helix in the alpha subunit (PDB accession code: 1PMA). Structural drawing produced using PyMOL (DeLano 2004).

The structural study of the proteasome revealed that it was a previously unclassified type of protease, which cleaved in an unspecific way to produce protein fragments with a range of six to nine amino acids. The protease activity was described as being an N-terminal nucleophile (Ntn) hydrolase (Brannigan *et al.* 1995).

In eukaryotes such as yeast and mammals, the core proteasomes are more complex than the simpler archaeal homologues and they have more than one type of α and β subunits and therefore a more complex structure (Groll *et al.* 1997; Unno *et al.* 2002). The 28

subunits of the yeast 20S proteasome are arranged as an ($\alpha1...\alpha7$, $\beta1...\beta7)_2$ complex in four stacked rings and occupy unique locations.

1.4.2 Assembly and activation of the proteasome

Five of the β type subunits of the yeast 20S proteasome are synthesised as proproteins before being protealytically processed for assembly into the particle. Of particular interest is that β1, β2 and β5 autolyse between the final residue in the prosequence (Gly-1) and the first residue in the active β subunit (Thr1) (Groll *et al.* 1997). With the generation of the N-terminus at Thr1 these three β subunits become active. The activity of these subunits was identified as follows: β5 carries the chymotrypsin-like (hydrophobic) and trypsin-like (basic) activity, β1 carries the peptidylglutamyl-peptide hydrolytic (PGPH) (acidic) activity, and β2 is suited for very large P1 residues of neutral, acidic or basic character. Only the β1, β2 and β5 subunits are proteolytically active, all other β subunits and all the α subunits are catalytically inactive.

The assembly of the proteasome is thought to start as an assembly of a precursor complex (called the 16S) consisting of α and unprocessed β subunits. The propeptides of the β subunits are thought to prevent the premature conversion of the precursor complex into matured, active, particles (Krüger *et al.* 2001). The proposed mechanism for the formation of the mature 20S proteasome is that during 16S precursor formation α subunits spontaneously oligomerise into seven-membered rings, onto which the β subunits assemble. After the formation of this 16S complex the β subunits must undergo a conformational change and rotate upwards, towards a second 16S complex undergoing the same conformational change. This rearrangement is induced either by the completed β ring or by the second 16S complex, possibly supported by chaperone proteins. After this conformational change has completed the β subunits from each 16S complex can bind to the other and so form an inactive 20S complex. The prosequence of the active β subunits would then be auto-catalytically cleaved to allow the activation of the complex (Mullapudi *et al.* 2004).

In the *T. acidophilum* proteasome structure entry ports with a diameter of ~13Å at each end of the cylindrical particles are open, however the N-terminal 12 residues are disordered in the structure. In the yeast 20S proteasome the N-termini of the subunits α1, α2, α3, α6 and α7 project into the opening and fill it completely with several layers of tightly interacting side-chains. This means that there can be no access for substrates

to the proteolytic chamber by this route without substantial conformational re-arrangement and thus, without activation, the cell is protected from the proteolytic activity of the proteasome.

Additional proteins are required to mediate proteasome function by overcoming the restrictions on substrate access imposed by the structure of the 20S proteasome and are therefore considered proteasome activators. The 20S proteasome forms complexes with different proteins and, thus, it is the centre of a modular system where different regulatory proteins confer unique catalytic and regulatory properties upon the proteasome. Of these proteasome activators the best studied are the 19S and 11S complexes and these will be discussed in turn here.

The asymmetric 19S activator binds to one or both of the α subunit disks of the 20S proteasome in a defined alignment and they face in opposite directions (Kopp and Kuehn 2003). When both ends of the 20S proteasome are bound to the 19S proteasome this complex is known as the 26S proteasome.

The 19S complex has a molecular weight of approximately 700 kDa and serves to recognise polyubiquitinated proteins and to convert them into a form suitable for degradation by the 20S core complex. Each 19S complex contains at least 17 different subunits and is assembled from two subcomplexes. An eight-subunit subcomplex of the 19S, the lid, can be dissociated from proteasomes prepared from a deletion mutant for Rpn10, a 19S subunit (Glickman *et al.* 1998). A second subcomplex, the base, contains all six proteasomal ATPases and links the 19S regulatory complex to the 20S proteasome. The base is sufficient to activate the 20S proteasome for degradation of peptides or a nonubiquitinated protein, whereas the lid is required for ubiquitin-dependent degradation (Glickman *et al.* 1998).

The yeast 19S subunit Rpn10 (S5a in humans) is involved in the recognition of polyubiquitinated substrates through its ubiquitin-interacting motif (UIM) (Deveraux *et al.* 1994; Fujiwara *et al.* 2004). The recognition of polyubiquitinated proteins is an important function as at this stage the specificity of the degradation of substrate proteins depends upon the correct recognition by the proteasome of targeted proteins. Rpn10 is not an essential gene in *S. cerevisiae* (Van Nocker *et al.* 1996) and so there is expected to be some redundancy in its function. This is surprising as the deletion of this subunit encourages the dissociation of the lid and base of the 19S complex. The ubiquitin

recognition function of the Rpn10 protein may be supplemented by the reversibly bound proteasome co-factor Rad23 that also recognises ubiquitin chains (Elsasser *et al.* 2004).

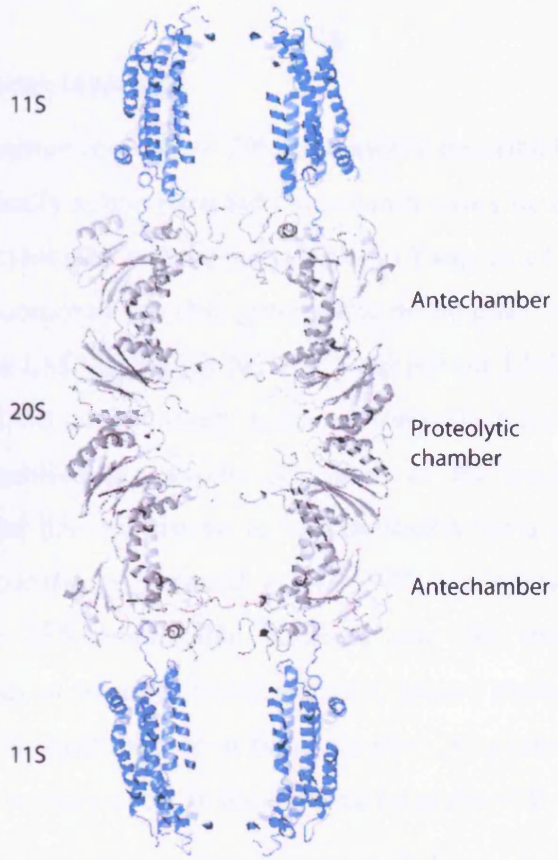

Figure 1.9 – 20S-11S proteasome complex

Ribbon representation of the structure of the 20S-11S complex (PDB accession code: 1FNT). The figure has been cut away to reveal the internal features and the gated access between the 11S and the antechamber of the 20S. 11S coloured in blue and 20S in grey. Figure produced using PyMol (DeLano 2004).

Another type of 20S proteasome activator is the non-ATP dependent activator, the 11S. This activator does not enable the proteasome to process proteins, only peptides, as it does not recognise ubiquitinated substrates. The crystal structure of the yeast 20S proteasome bound to the 11S activator complex from *Trypanosoma brucei* showed that the 11S activation loops induces conformational changes in the α subunits of the 20S proteasome that open the gate separating the proteasome interior from the intracellular environment (Whitby *et al.* 2000). In addition to this the activation loops also impose a more stringent sevenfold symmetry on the 20S proteasome thereby straightening the asymmetrically oriented tails of the α subunits and moving them away from the entrance/exit gates (Förster *et al.* 2003). This ordered open conformation is required for

proteolysis and its formation is likely to be triggered by outward displacement of the α subunit residues in all activators, both ATP-dependent and -independent (Förster *et al.* 2005).

1.4.3 The immunoproteasome

In addition to the constitutively active 20S proteasome described above, in mammalian cells the three catalytically active proteasomal subunits many be further regulated by the immunomodulatory cytokine, interferon-γ (IFN-γ) (Yang *et al.* 1992). These IFN-γ induced β subunits incorporate into the proteasome in the place of the active β subunits and are termed β1i (or LMP2), β2i (or MECL1) and β5i (or LMP7) and replace the β1, β2 and β5 subunits of the constitutively active proteasome in a controlled way (Früh *et al.* 1994). The modified proteasome is known as the immunoproteasome. The introduction of β1i and β5i was shown to be responsible for a change in the cleavage specificity of the proteasome (Driscoll *et al.* 1993; Gaczynska *et al.* 1993). The incorporation of the IFN-γ inducible subunits into the proteasome favours the production of the types of peptides found on MHC class I molecules, which terminate almost exclusively with hydrophobic or basic residues (Gaczynska *et al.* 1993) and so the immunoproteasome plays a role in the immune response of the cell.

The structure of the mammalian proteasome allowed the prediction of the structure of the immunoproteasome (Unno *et al.* 2002). The sequence identities of the human β1i, β2i and β5i subunits to the human β1, β2 and β5 are 59.2%, 57.7% and 68.6% respectively. The remaining 11 subunits are not changed between the proteasome and immunoproteasome. In the predicted immunoproteasome structure the S1 pocket of β1i was more apolar than that of β1, making it likely that the PGPH activity of this subunit would be reduced and chymotrypsin-like activity enhanced. There was no large structural change predicted for the β2 replacement with β2i. This change fits with the observed change in specificities for the cleavage of the polypeptides.

The formation of the immunoproteasome must occur in such a way as to prevent the formation of hybrid proteasomes that contain both the constitutively active β subunits and the inducible immuno β subunits. During immunoproteasome formation the incorporation and processing of inducible proteasome β subunits are cooperative processes. Subunit β1i is essential for the incorporation of β2i. The contribution of β5i appears to be to accelerate the β1i and β2i processing kinetics. This cooperative

behaviour is not dependent on the active sites of the subunits and has the end result of creating a homogenous proteasome population (Schmidt *et al.* 1999).

The proteasome plays a vital role in our immune system by generating peptides from intracellular antigens that are then presented to T cells at the cell surface. In response to viral attack the specially adapted immunoproteasome is a more efficient protease for the production of the peptides for presentation by the MHC class I molecules than the constitutively active proteasome (Kloetzel 2004). Bioinformatic study showed that the immunoproteasome is a more specific enzyme than the constitutively active proteasome and produces better ligands for MHC class I presentation when analysing the degradation of pathogen proteomes (Kesmir *et al.* 2003). The MHC class I presentation pathway enables T cells to distinguish between intracellular self-antigens and cellular pathogens. It is likely that the immunoproteasome has co-evolved with the major histocompatibility complex to optimise antigen presentation (Kesmir *et al.* 2003).

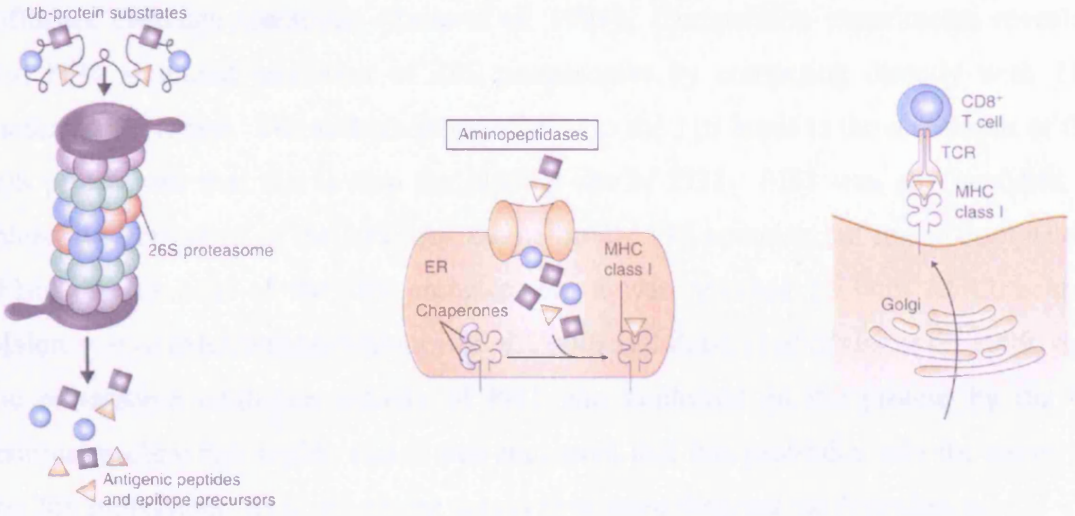

Figure 1.10 – The proteasome and antigen processing

Substrates that contain MHC class I epitopes are tagged with a polyubiquitin chain and targeted to the proteasome for destruction. Proteasomal processing results in the generation of small peptides that are transported to the ER by TAP. For N-terminal extensions the peptides may be trimmed in the cytosol or ER by aminopeptidases. In the ER peptides bind and stabilise the appropriate MHC class I molecule with the help of chaperones. Fully loaded and assembled MHC class I molecules are translocated to the cell surface. By binding specifically to an MHC class I molecule loaded with a unique peptide, cytotoxic $CD8^+$ T cells with complementary T cell receptors (TCRs) are stimulated to proliferate and destroy infected cells. Figure reproduced from (Kloetzel 2004).

1.4.4 Proteasome inhibitor protein 31, PI31

In addition to activators of the proteasome natural inhibitors have also been identified. PI31 was first identified as an inhibitor of the proteasome after purification from bovine red blood cells (Chu-Ping *et al.* 1992). PI31 was found to be a protein of about 31kDa, which may be present in cells in different multimeric forms and was shown *in vitro* to be a specific non-competitive inhibitor of the 20S proteasome but not of the 26S proteasome.

Studies of recombinant PI31 in the Sijts lab (Zaiss *et al.* 1999) and the DeMartino lab (McCutchen-Maloney *et al.* 2000) revealed more of the function of this protein. PI31 is a proline rich protein, particularly at its C-terminus where 26% of the amino acids are prolines (McCutchen-Maloney *et al.* 2000). PI31 was shown to inhibit both the constitutively active 20S proteasome and the immunoproteasome with the same kinetics when analysing the degradation of short fluorogenic peptides *in vitro* and it does not influence cleavage specificity (Zaiss *et al.* 1999). Competition experiments revealed that PI31 mediated inhibition of 20S proteasomes by competing directly with 11S mediated activation. The authors proposed that as the 11S binds to the α subunits of the 20S proteasome that this is also the binding site of PI31. PI31 was also reported to inhibit the activation of the 20S proteasome by the 19S complex but not to completely ablate the function of the 26S proteasome. It was reported by both McCuthchen-Maloney *et al* (McCutchen-Maloney *et al.* 2000) and Zaiss *et al* (Zaiss *et al.* 1999) that the proteasome inhibition activity of PI31 was conferred on the protein by the C-terminal proline rich region and it was suggested that this protrudes into the cavity of the 20S proteasome, thus preventing access to proteins targeted for degradation.

Further experiments carried out in the Sijts lab revealed that PI31 induced inhibition of the proteasome could not be observed *in vivo* (Zaiss *et al.* 2002) and so another role for PI31 was sought. Over expression of PI31 in mouse embryonic cells has no impact on either proteasome mediated proteolysis or upon cell cycle progression. Instead PI31, when up regulated, interferes with the maturation of the immunoproteasome. Over expression of PI31 abrogates MHC class I presentation of an immunoproteasome dependent cytotoxic T lymphocyte epitope and reduces the surface MHC class I levels on IFN-γ treated mouse embryonic cells in culture. The authors suggested that the function of PI31 may be that of interference with the processing of the prosequences of

the IFN-γ inducible subunits thus preventing the formation of fully mature immunoproteasomes.

In addition to its role interacting with the proteasome PI31 has also been linked to CD2BP2 protein (Kofler *et al.* In Press), which is a cellular adapter protein that was originally identified as a binding partner of the T cell adhesion protein CD2 in the context of T cell signalling. A small α/β protein, the GYF domain of the protein CD2BP2, serves as an adapter that recognizes proline-rich sequences in intracellular proteins. This GYF domain has been shown to interact with a C-terminal motif in PI31 by spot analysis and yeast two-hybrid (Kofler *et al.* In Press). PI31 contains both types of binding motifs identified to be the consensus motifs required for interaction with the GYF domain that is; PPG[W,F,Y,M,L] and PPG*x*[R,K]. The implications of this new interaction on the function of PI31 have yet to be uncovered and indeed this interaction has not yet been demonstrated *in vivo*.

The successful use of synthetic proteasome inhibitors in the treatment of cancer implies that the naturally occurring PI31 may be amenable to manipulation for the same purpose. This may warrant further investigation into the function of this protein *in vivo*.

1.5 Aims and objectives of this thesis

- Structural characterisation of the Fbxo7-Skp1 complex, its homologues and interaction partners.
- Identification of the substrates and binding partners for SCF^{Fbxo7} and validation of the interaction.

By investigating the F-box protein Fbxo7 I hoped to employ several strategies to learn more about the structure of Fbxo7 and its interaction with Skp1. I also hoped to undertake biochemical screens to find interacting proteins and therefore potential substrates. These proteins could, in turn, provide some insight into the molecular and cellular function of Fbxo7.

Ultimately I hoped to reconstitute the SCF^{Fbxo7} E3 ligase *in vitro* and determine its 3D structure.

2 Preparation and characterisation of the Fbxo7-Skp1 complex and its binding partners

The protein Fbxo7 has been shown to be a proto-oncogene (Laman *et al.* 2005) but much of its structure and function remain uncharacterised. This chapter describes the bioinformatic analysis of Fbxo7 leading to the identification of previously unidentified domains within this protein. These domain boundaries were confirmed by limited proteolysis and used to direct the production of recombinant Fbxo7-Skp1 complex for crystallisation trials. *In vivo* and *in vitro* characterisation of Fbxo7 revealed that it is a subunit of an SCF complex and identified a novel binding partner for Fbxo7. A flowchart describing the experiments performed can be found below (Figure 2.1).

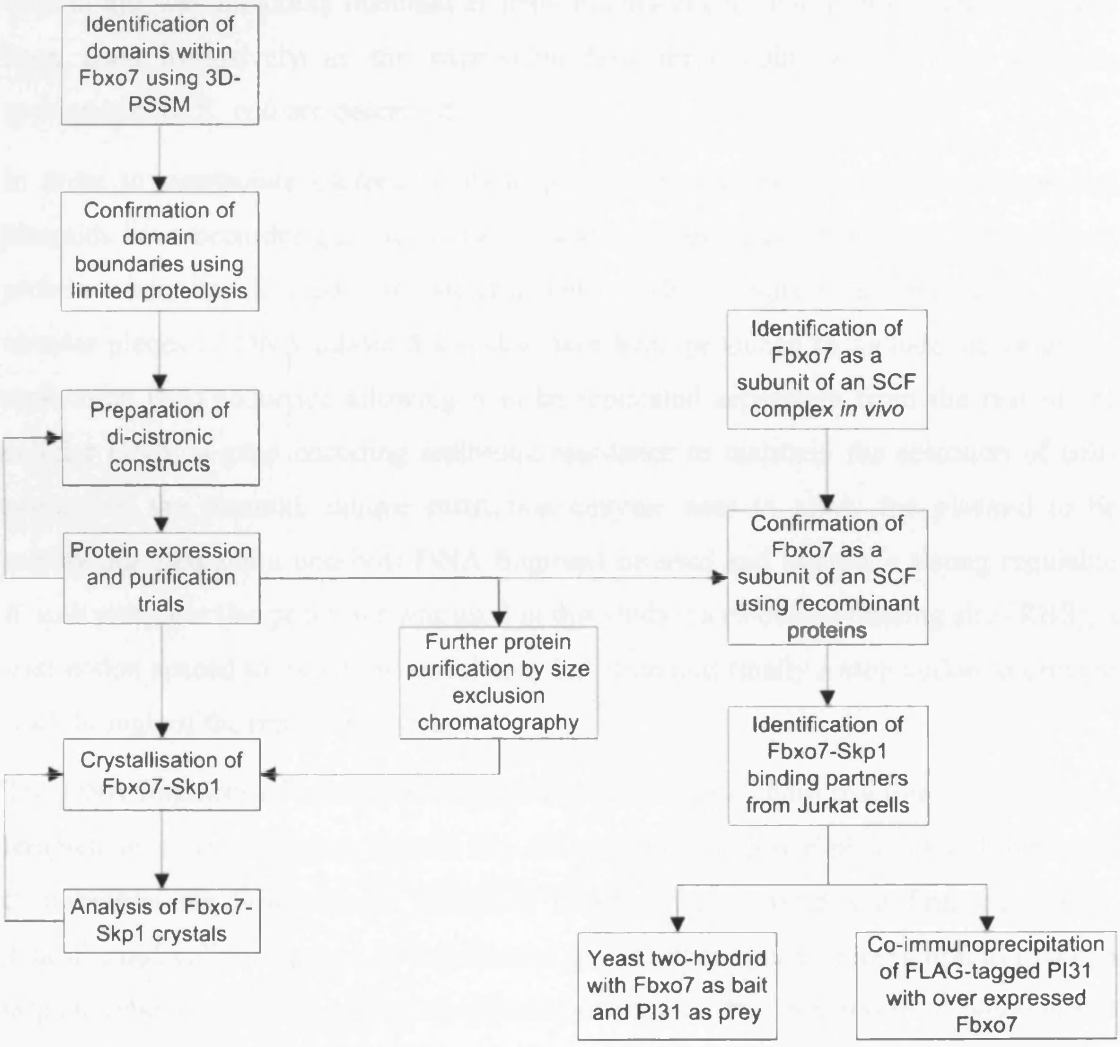

Figure 2.1 – Flowchart of the experiments described in chapter two.

2.1 Preparation of recombinant Fbxo7-Skp1 complexes

2.1.1 General aspects of recombinant protein production

Many biologically interesting proteins are present in such low quantities under endogenous conditions within cells that extracting sufficient quantities of these proteins for analysis from their naturally producing organisms would be difficult. Recombinant DNA techniques originally developed in the 1970s by Cohen *et al* (Cohen *et al.* 1973) and developed since have helped to overcome this problem. Our increased understanding of the molecular mechanisms of gene expression in bacteria has enabled the manipulation of this system to express foreign proteins from cloned genes at high levels (Appelbaum and Shatzman 1999). Several different types of host cells can be used in this way including mammalian cells and insect cells but in this study *E. coli* has been used exclusively as the expression host throughout and so only methods appropriate for *E. coli* are described.

In order to manipulate bacterial cells to produce human proteins extra-chromosomal plasmids have been designed to transcribe and translate large amounts of heterologous protein when transformed into bacterial cells. These expression vectors are small circular pieces of DNA (about 5 kb) that have been produced to include; an origin of replication (ori) sequence allowing it to be replicated separately from the rest of the cellular DNA; a gene encoding antibiotic resistance to maintain the selection of cells containing the plasmid; unique restriction enzyme sites to allow the plasmid to be readily digested and a non-host DNA fragment inserted and ligated; a strong regulable *E. coli* promoter (lac promoter was used in this study); a ribosome binding site (RBS); a start codon spaced to maximise translation initiation and finally a stop codon to prevent read through of the remaining vector.

The DNA fragment of interest is amplified by polymerase chain reaction (PCR) from a template or a cDNA library, ligated into the previously digested plasmid and then used to transform the host cell (as described in appendix A) (Old and Primrose 1994). Transformed cells are grown to exponential growth phase prior to induction to obtain a large number of cells containing the plasmid of choice. The expression of proteins that can be toxic to the bacteria is controlled using the bacterial lac operator. The lac repressor is bound under normal growth conditions and prevents transcription.

Induction of expression is achieved by the addition of Isopropyl-β-d-Thiogalacto-Pyranoside (IPTG), which binds to the lac repressor, which thus dissociates from the operator allowing translation to proceed (Appelbaum and Shatzman 1999).

To facilitate rapid purification of recombinant protein produced in this way a variety of tags are employed. These tags consist either of peptides, such as a poly-histidine (His) tag, or of globular proteins, such as glutathione S-transferase (GST) and maltose binding protein. These tags bind to small molecule ligands, which can be immobilised on Sepharose beads to form an affinity column and can then be used to separate the tagged recombinant proteins from the bacterial proteins under non-denaturing conditions. Recombinant proteins are readily eluted from the affinity column by the addition of a site specific protease site between the tag and protein of interest. A method originally described by Smith and Johnson (Smith and Johnson 1988) for expression of all GST fusion proteins was used in this thesis. This method was improved by Guan and Dixon (Guan and Dixon 1991) by the addition to the original vector of a poly-glycine linker inserted after the thrombin cleavage site, which improved cleavage efficiency. The linker allows cleavage while attached to the affinity resin and therefore efficient single step purification. The purification procedure outlined by Guan and Dixon (Guan and Dixon 1991) and described in appendix A was employed for all GST fusion proteins described here.

This method for producing recombinant protein allows the adaptation of the proteins to contain the optimal domains for either structural or functional characterisation. For functional characterisation it is preferable to analyse the full length protein however in crystallography the identification and removal of flexible regions of proteins can greatly improve the quality of crystals. Using secondary structure prediction servers to identify structured domains within a protein can identify flexible regions and limited proteolysis on full length proteins can also identify the minimal structured domain of the protein.

2.1.2 Identification of domains within Fbxo7

Identification of the regions of Fbxo7 that are likely to be globular domains and therefore represent good crystallisation targets was performed by domain analysis. The biomolecular modelling server 3D-PSSM (3 dimensional position specific scoring matrix) (Kelley A.L. 2000) was used to predict the secondary structure of Fbxo7 and to identify domains with remote sequence similarity which are below the threshold of

sequence alignment servers. This analysis revealed an N-terminal ubiquitin like domain (Ubl) spanning amino acids 1-78 with an E-value of 1.6 and an estimated precision of 65% (Figure 2.2a). The F-box domain itself was shown to span residues 329-375 with an E-value of 2 and an estimated precision of 60% (Figure 2.2c). A further unknown globular domain was identified preceding the F-box domain with a predicted α/β structure but with no known function (Figure 2.2b). This domain spanned residues 173-324. 3D-PSSM also revealed a predicted unstructured C-terminal after the F-box domain, which on analysis of the sequence was revealed to be a proline-rich region (PRR) and would be unlikely to crystallise. The domains identified in this search are outlined in Figure 2.2 with the sequence alignment of the different domains within Fbxo7 proteins from all species. The secondary structure predicted from 3D-PSSM shows that the Ubl domain is predicted to have a α/β fold and the F-box domain is predicted to contain three helices, consistent with previous structural information (see chapter one for details of these domains).

Having identified the likely domain boundaries of Fbxo7 using bioinformatics Phillip Knowles (Structural Biology Laboratory, Cancer Research UK) carried out limited proteolysis on Fbxo7-Skp1 (Koth *et al.* 2003) (Figure 2.3). Recombinant Fbxo7(129-398)-Skp1(1-146) was used because Fbxo7(129-398) had been shown by Laman *et al* (Laman *et al.* 2005) to interact with viral cyclin by yeast two-hybrid and so it is likely that Fbxo7(129-398) represents a functionally relevant domain. Limited proteolysis can be used to reveal the areas of a protein that are more susceptible to proteolysis and are therefore more likely to be flexible or exposed regions of the protein. Structured regions are less susceptible to proteolysis and therefore would not be degraded by this method and so can be identified. The analysis of the protein fragments produced by limited proteolysis was performed by N-terminal Edman sequencing by Nick Totty (Protein Analysis Laboratory, Cancer Research UK). Fbxo7(131-C-terminal truncation) and Fbxo7(171-398) were revealed to be likely to have defined secondary structure by this method. The latter correlates well with the results from the bioinformatic analysis, amino acids 171-398 include both the novel domain and the F-box domain. Skp1(1-146) shows no sign of proteolysis and therefore is likely to be a globular domain.

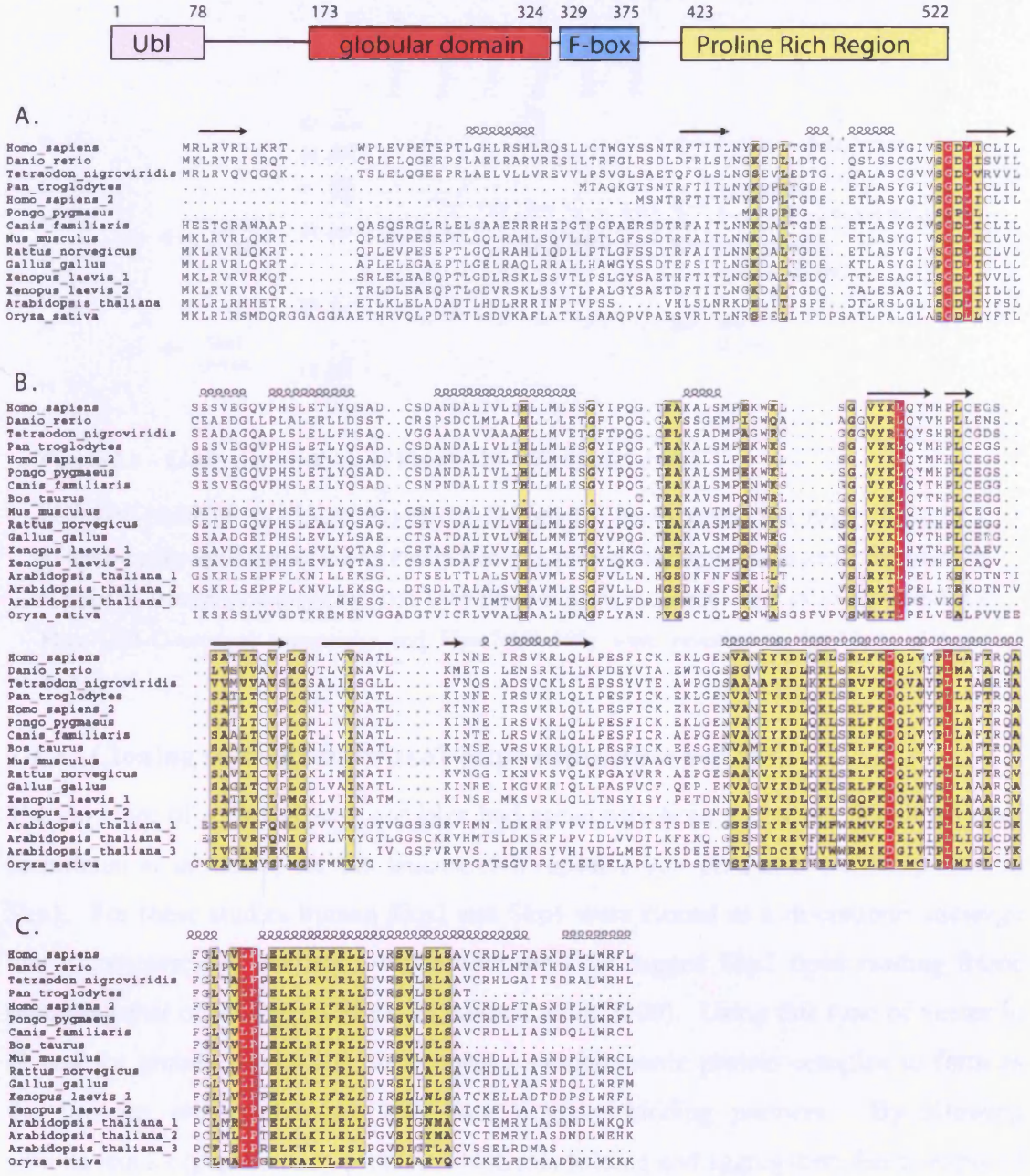

Figure 2.2 – Sequence alignment of Fbxo7 domains identified by 3D-PSSM

3D-PSSM (Kelley A.L. 2000) was used to identify all globular domains within Fbxo7. The secondary structure prediction revealed both the Ubl domain and the novel domain to have a predicted α/β fold. The F-box domain is likely to be α helical. Sequence alignments of Fbxo7: A) Ubl domain, B) Uncharacterised globular domain, C) F-box domain. Sequence alignment was performed using ClustalX (Thompson *et al.* 1997) and figure produced using ESPript (Gouet *et al.* 1999).

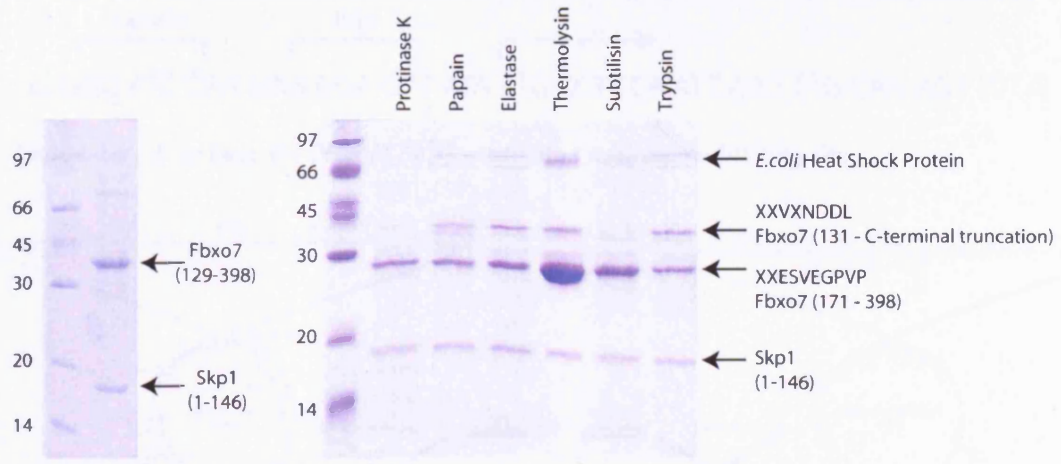

Figure 2.3 – Limited proteolysis of Fbxo7(129-398)-Skp1(1-146)

Left: SDS-PAGE of Fbxo7(129-398)-Skp1(1-146) prior to limited proteolysis. Right: SDS-PAGE of limited proteolysis carried out by Phillip Knowles, protein fragments and sequence identified by N-terminal Edman sequencing by Nick Totty (Protein Analysis Laboratory, Cancer Research UK). Fbxo7(131-C-terminal truncation) and Fbxo7(171-398) were revealed to be likely globular domains.

2.1.3 Cloning strategy for Fbxo7-Skp1 expression vectors.

The structure of a Skp1-F-box complex had previously been solved by Schulman *et al* (Schulman *et al.* 2000) for the leucine rich repeat F-box protein Skp2 complexed to Skp1. For these studies human Skp2 and Skp1 were cloned as a di-cistronic message into a commercially available vector with the GST tagged Skp2 open reading frame preceding that of un-tagged Skp1 (Schulman *et al.* 2000). Using this type of vector in *E. coli* for protein production encourages a stoichiometric protein complex to form as the proteins are synthesised with their cognate binding partners. By allowing simultaneous expression any problems with mis-folding and aggregation due to exposed hydrophobic contact interfaces are reduced due to the presence of the physiological binding partner. As only one of the proteins was expressed as a GST fusion protein, purification by affinity chromatography allows for the detection of a properly folded partner protein.

To produce sufficient recombinant Fbxo7 for crystallisation we co-expressed Fbxo7 with Skp1, its *in vivo* partner, in order to mimic *in vivo* conditions and enhance the stability of Fbxo7. Many F-box proteins have been shown to interact directly with Skp1 (Bai *et al.* 1996) and Fbxo7 in particular was shown to bind to Skp1 *in vivo* in HeLa cells (Cenciarelli *et al.* 1999).

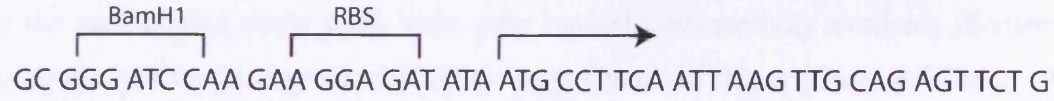

Figure 2.4 – 5' primer for PCR of Skp1 containing a ribosome-binding site

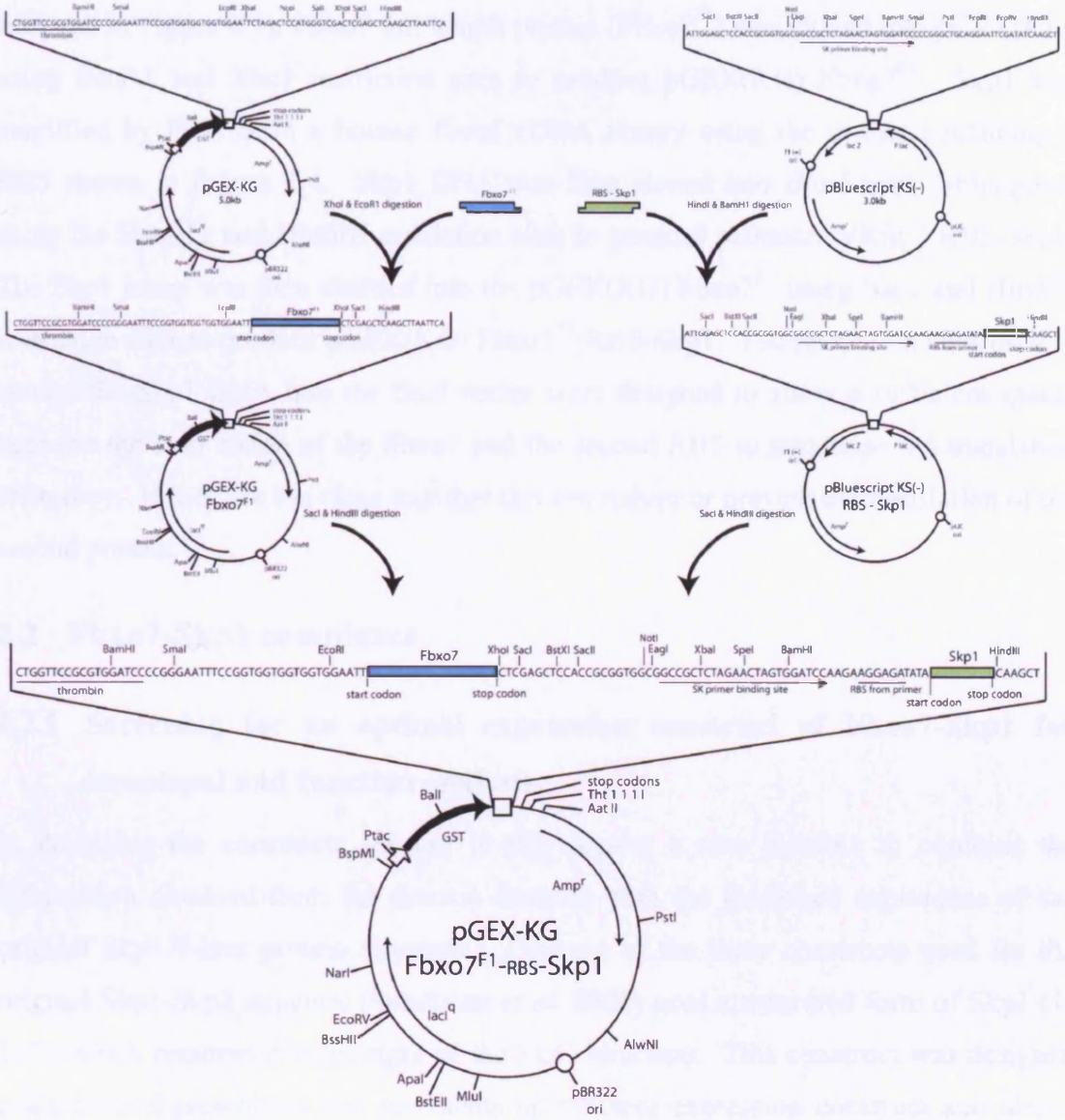

Figure 2.5 – Construction of expression plasmids

Fbxo7 (blue bar) was inserted into pGEX(KG) and Skp1 (green bar) was inserted into pBluescript. Fragments from these two plasmids were then ligated to produce pGEX(KG) Fbxo7^{F1}-RBS-Skp1. Constructs were produced by Heike Laman. The thick filled arrow (➡) represents the coding sequence for GST. Unique restriction sites are indicated. The locations of the genes for ampicilin resistance (Amp^r) and the lac repressors (lacI^q, lac Z and P lac) are indicated by thin arrows (→).

At the start of this study there were only limited commercially available di-cistronic constructs and so we adapted pGEX(KG) from Guan and Dixon (Guan and Dixon 1991) to contain an additional RBS. Heike Laman (Wolfson Institute, UCL) produced all Fbxo7 - Skp1 constructs using the following procedure (outlined for full length Fbxo7 here and in Figure 2.5); Fbxo7 full length protein ($Fbxo7^{F1}$) was cloned into pGEX(KG) using EcoR1 and Xho1 restriction sites to produce pGEX(KG) $Fbxo7^{F1}$. Skp1 was amplified by PCR from a human foetal cDNA library using the primer containing a RBS shown in Figure 2.4. Skp1 DNA was then cloned into BlueScript (Stratagene) using the BamH1 and HindIII restriction sites to produce pBluescriptKS(-) RBS-Skp1. The Skp1 insert was then shuttled into the pGEX(KG) $Fbxo7^{F1}$ using Sac1 and HindIII restriction sites to produce pGEX(KG) $Fbxo7^{F1}$-RBS-Skp1. The restriction sites used to shuttle the Skp1 DNA into the final vector were designed to allow a sufficient spacer between the stop codon of the Fbxo7 and the second RBS to maximise the translation efficiency. If they are too close together this can reduce or prevent the translation of the second protein.

2.2 Fbxo7-Skp1 complexes

2.2.1 Screening for an optimal expression construct of Fbxo7-Skp1 for structural and function analysis

In designing the constructs for use in this project it was possible to combine the information obtained from the domain analysis with the published experience of the original Skp1-F-box protein structure. Two out of the three constructs used for the original Skp1-Skp2 structure (Schulman *et al.* 2000) used a truncated form of Skp1 (1-147), which removes α helix eight of the Skp1 structure. This construct was designed using limited proteolysis and represents an efficient expression construct and also a good crystallisation target. In all the constructs of Fbxo7-Skp1 produced for this project the truncated form of Skp1 (1-146) has been used and will be referred to as Skp1ΔH8.

Eleven different constructs of human Fbxo7-Skp1ΔH8 were produced, sequenced and expression trials were performed in *E. coli* under standard protocol conditions, outlined in appendix A (Figure 2.6 and Figure 2.7). Throughout this chapter all constructs will be referred to using the labelling assigned in Figure 2.6 and Figure 2.7 as a superscript e.g. $Fbxo7^{F1}$ - Skp1.

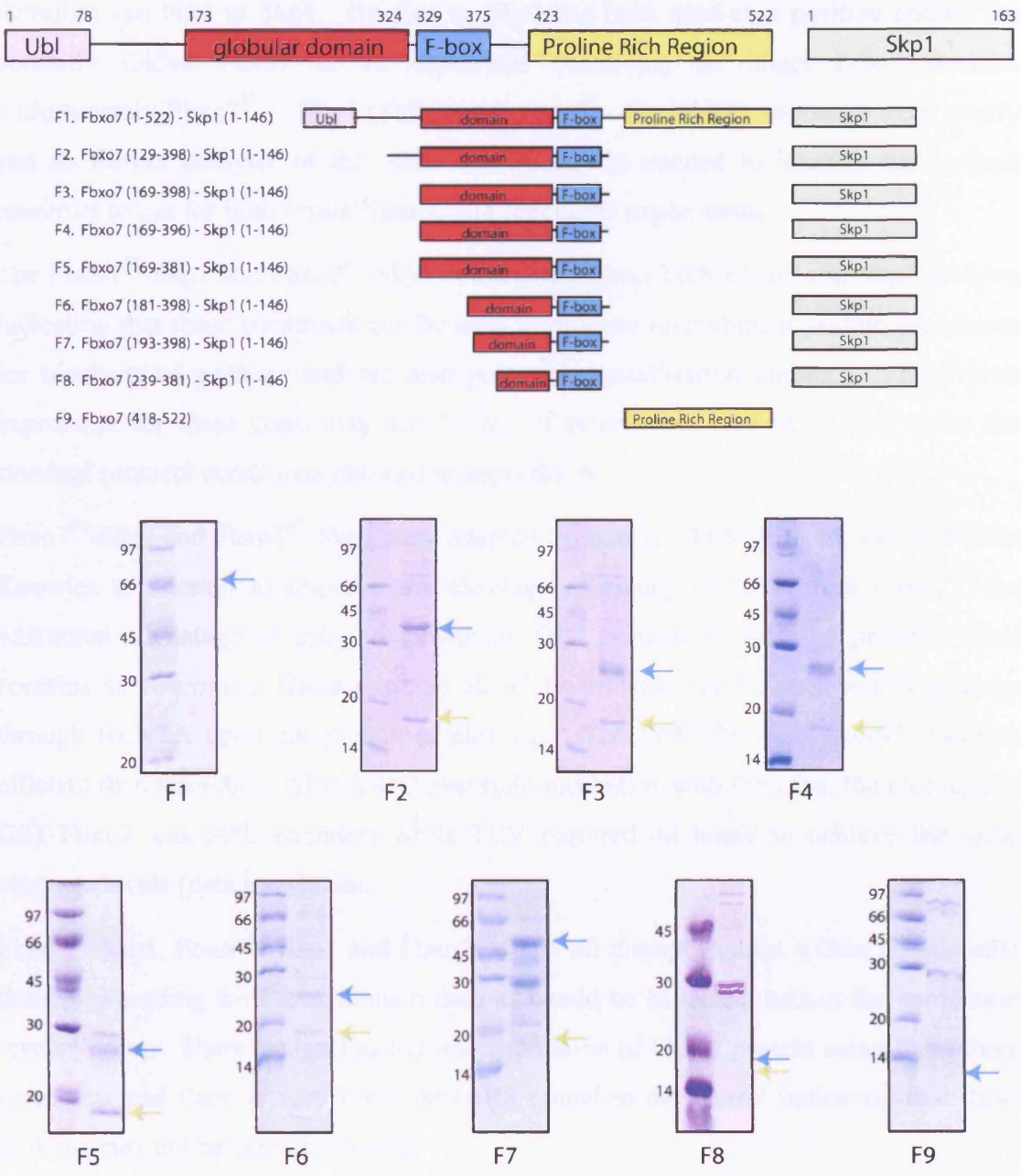

Figure 2.6 – Constructs produced and expression trials for Fbxo7-Skp1 complexes

Expected positions of Fbxo7 (blue) and Skp1ΔH8 (green) generated by the expression constructs shown are marked with arrows. Fbxo7^{F7} is shown as the un-cleaved GST fusion protein. F1 produced very little detectable protein, SDS-PAGE provided by Phillip Knowles, and neither F8 nor F9 produced any detectable protein. F4 and F5 have both been adapted to contain a TEV protease site by Phillip Knowles and this has been used for cleavage of the GST moiety. Skp1 is not a fusion protein and so was only detectable when Fbxo7 binds correctly to it indicating that both proteins had folded correctly.

All of the constructs produced except Fbxo7^{F9} contain the F-box domain of Fbxo7 and therefore can bind to Skp1. Binding to Skp1 has been used as a positive control for correctly folded Fbxo7 in all constructs containing an intact F-box domain. Unfortunately Fbxo7^{F1} – Skp1 (Full length Fbxo7 – Skp1ΔH8) expresses very poorly and so further analysis of the other constructs was needed to identify the optimal construct to use for both crystallisation and functional experiments.

The Fbxo7^{F2}-Skp1 and Fbxo7^{F3}-Skp1 constructs express both Fbxo7 and Skp1 proteins indicating that these constructs can be used to produce recombinant protein complexes for biochemical analysis and are also potential crystallisation targets. The overall expression for these constructs was 1-2mg of protein per litre of *E. coli* under the standard protocol conditions outlined in appendix A.

Fbxo7^{F4}-Skp1 and Fbxo7^{F5}-Skp1 were adapted to contain a TEV protease site by Phillip Knowles to attempt to improve the cleavage efficiency of GST from Fbxo7. An additional advantage of using recombinant TEV protease is that the protease itself contains an N-terminal His-tag and so all of the protease can be removed by passage through Ni-NTA upon completion of cleavage. The TEV cleavage of GST was less efficient than thrombin. After a 4°C overnight incubation with thrombin the cleavage of GST-Fbxo7 was 80% complete while TEV required 48 hours to achieve the same cleavage levels (data not shown).

Fbxo7^{F6}-Skp1, Fbxo7^{F7}-Skp1 and Fbxo7^{F8}-Skp1 all disrupt regions within the globular domain preceding the F-box domain that, as would be expected, causes the expression levels to drop. There is significantly less expression of Fbxo7 protein using these three constructs and there is very little Skp1ΔH8 bound to the Fbxo7 indicating that these proteins may not be correctly folded.

The constructs previously described only varied the Fbxo7 sequence and Skp1ΔH8 was kept constant. It was also important to identify the optimum Skp1 construct to use for efficient translation. When Schulman *et al* (Schulman *et al.* 2000) identified Skp1ΔH8 as an appropriate construct they also performed two loop deletions in their construct design. They solved three versions of the Skp1-F-box protein complex and all three of the constructs that were used to produce the proteins lack a six residue internal loop of Skp1 (38-43). This region was determined to be unstructured on the basis of limited proteolysis and is absent from several Skp1 orthologues. One of the three structures

solved also lacked a protease sensitive loop (71-82), which was disordered in the other two Skp1 crystal structures. After identification of the optimal region of Fbxo7 for expression as amino acids 169 to 398 (Fbxo7^{F3}) this construct was adapted to remove the two loop regions identified by Schulman *et al* (Schulman *et al.* 2000) (Figure 2.7).

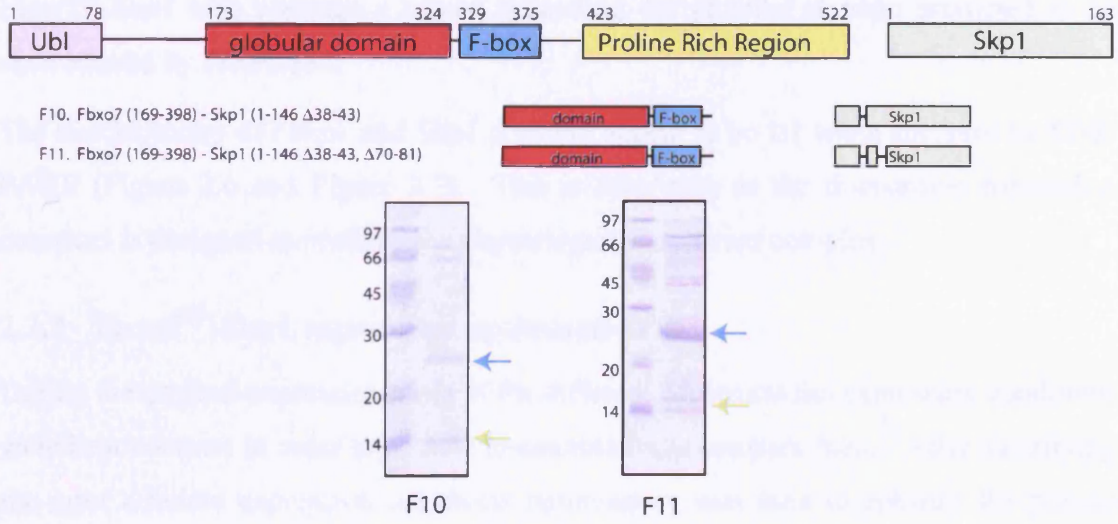

Figure 2.7 – Fbxo7^{F3}-Skp1 construct adapted to contain loop deletions of Skp1

Expected positions of Fbxo7 (blue) and Skp1ΔH8 (green) generated by the expression constructs shown are marked with arrows. F10 shows low expression and F11 Skp1 appears to show evidence of proteolytic clipping. Fbxo7^{F3}-Skp1 was adapted to contain the Skp1 loop deletions by Heike Laman.

Fbxo7-Skp1^{F10} contains the 38-43 loop deletion used in all the previous Skp1 structures. Skp1 protein is visible on SDS-PAGE indicating that this deletion does not interfere with the binding between Skp1 and the F-box domain of Fbxo7. As this loop deletion of Skp1 had been used successfully in three other structure determinations it was unfortunate that the expression of Fbxo7-Skp1^{F10} was not high enough to continue to use this construct for crystal trials.

Fbxo7-Skp1^{F11} contains the double loop deletion used in the Skp1-Skp2 structure for one of their datasets (38-43 and 71-82) (Schulman *et al.* 2000). The expression of the protein complex from this construct appears to produce two discrete Skp1 bands although these have not been identified by mass spectrometry or Western-blot analysis. The expression of this construct was again too low to follow up in our crystal trials at less than 0.5mg/l.

On completion of all the initial construct expression trials two constructs were revealed to produce reasonable yields of protein complex, Fbxo7^{F2}-Skp1 and Fbxo7^{F3}-Skp1. The two complexes both contain the full globular domain immediately preceding the F-box and the complete F-box domain with some residues after the C-terminus of the F-box. Fbxo7^{F2}-Skp1 also contains a region preceding the globular domain predicted to be unstructured by 3D-PSSM.

The stoichiometry of Fbxo7 and Skp1 does not appear to be 1:1 when analysed by SDS-PAGE (Figure 2.6 and Figure 2.7). This is surprising as the di-cistronic expression construct is designed to produce the physiologically relevant complex.

2.2.2 Fbxo7^{F3}-Skp1 expression optimisation

During the original expression trials of the different constructs the expression conditions were kept constant in order to be able to quantitatively compare them. After identifying the most efficient expression constructs optimisation was used to enhance the protein yield.

Variables that can be used to improve protein expression in *E. coli* include the type of media used, temperature of cell growth and induction, and the concentration of IPTG used for protein expression induction. An expression trial was performed varying each of these variables in turn. Trials were performed at an induction temperature of 18°C, 25°C and 37°C using LB, TB and NZY media. IPTG concentration was varied dependent on the temperature, 50μM and 100μM final concentration used at 18°C and 100μM, 250μM and 500μM final concentration used at both other temperatures (Figure 2.8).

The expression trials produced high yields of Fbxo7 in all conditions tested; unfortunately this was not matched by Skp1ΔH8 expression. This is surprising as the stoichiometry of the proteins would be expected to be 1:1 due to their relationship as an obligate heterodimer. The yields of Skp1ΔH8 were dependant on all three of the variables tested as can be seen in Figure 2.8. There was a reasonable yield of Skp1ΔH8 using TB media at both concentrations of IPTG at 18°C while using either LB or NZY medias under any conditions there were considerably reduced yields. Growing up *E. coli* using TB media generally produces a larger cell pellet than LB media, as it is a richer food source, which could account for some of the difference in expression seen.

There was a higher protein complex yield using TB media at all temperatures tested but the most efficient production of protein occurs using TB as the media at 18°C with an IPTG concentration of 50μM. This method was used alongside the LB protein production described in appendix A to produce protein complexes for experimental analysis.

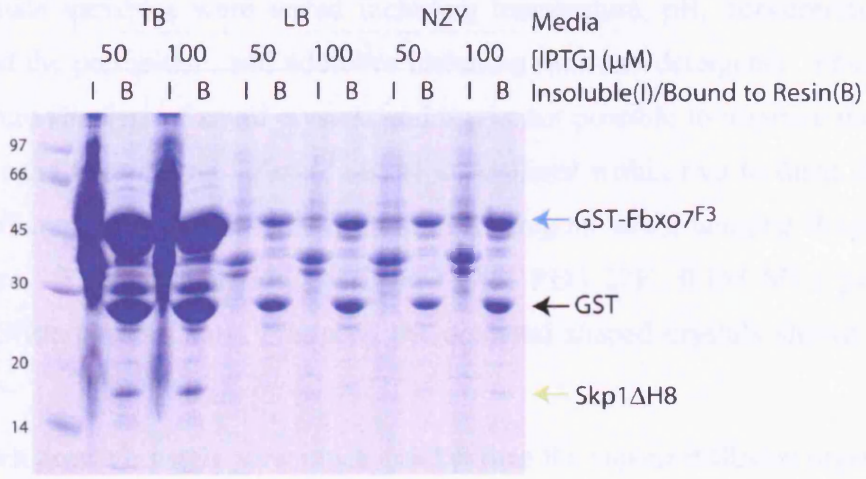

Figure 2.8 – Expression trials of Fbxo7^{F3}-Skp1

Typical SDS-PAGE of an expression trial showing three different media tested with varying concentrations of IPTG at 18°C. All lanes other than those using TB media at 18°C contained similar amounts of Fbxo7^{F3}-Skp1 (other temperatures not shown). Expected positions of expressed proteins are marked with arrows, GST-Fbxo7^{F3} fusion protein (blue), Skp1ΔH8 (green) and free GST (black). The insoluble lanes (I) represent equal volumes of insoluble *E. coli* cell pellet and the bound to resin lanes (B) represent equal volumes of purified protein bound to glutathione Sepharose affinity resin.

2.3 Crystallisation of Fbxo7-Skp1 complexes

Fbxo7^{F2}-Skp1 and Fbxo7^{F3}-Skp1 were both submitted to crystal trials by Philip Knowles and myself. Initial trials were performed using both vapour diffusion and microbatch methods as described in appendix B. There is no known *a priori* way of predicting which condition any given protein or complex will crystallise in. Targeted trial and error experiments were performed using sparse matrix screens based on the conditions used previously to successfully crystallise proteins.

Crystals were obtained from both Fbxo7^{F2}-Skp1 and Fbxo7^{F3}-Skp1 after considerable screening of conditions. The initial trials used commercially available sparse matrix screens from Hampton (Jancarik and Kim 1991; Cudney *et al.* 1994) and Emerald

BioSystems and a protein complex screen made using the instructions in Radaev and Sun (Radaev and Sun 2002). These screens gave an indication of which conditions to follow up with grid screens and additive screens. Crystals of Fbxo7^{F2}-Skp1 originally formed in Hampton crystal screen 2 condition 22 and those of Fbxo7^{F3}-Skp1 formed in Hampton crystal screen 2 in conditions 35 and 39. To improve the quality and size of these crystals variables were tested including temperature, pH, concentration of the protein and the precipitant, and additives including salts and detergents. Fbxo7^{F2}-Skp1 crystals were showers of small crystals and it was not possible to improve their quality using the conditions tested. Fbxo7^{F3}-Skp1 crystallised within two to three days under vapour diffusion at a protein concentration of 16mg/ml using hanging drops at room temperature. The well solution consisted of 6% PEG 20K, 0.1M MES pH 6.5 and 0.005% Zwittergen 3.12 and produced the diamond shaped crystals shown in Figure 2.9a.

Micro-batch grown crystals grew much quicker than the vapour diffusion crystals taking only 15 minutes to appear and did not require the addition of detergent to form ordered crystals. It was not possible to produce large enough crystals to handle using this method.

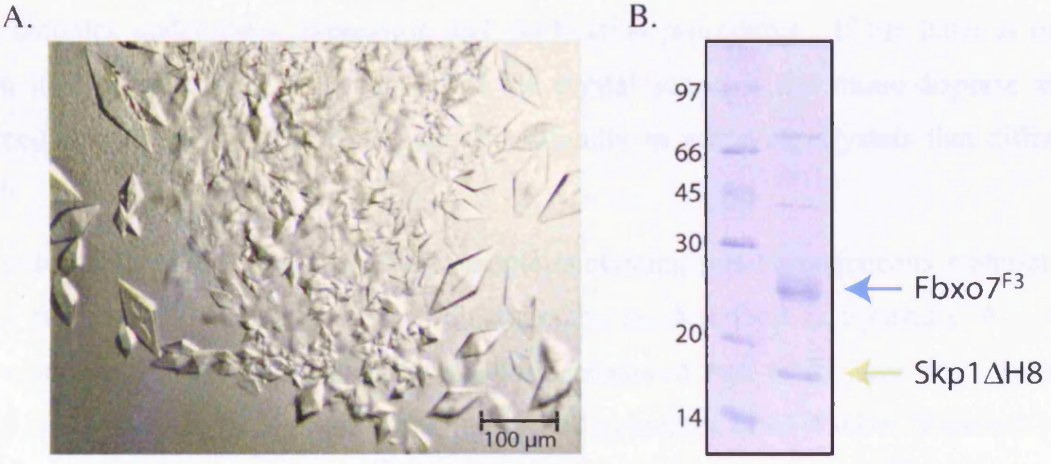

Figure 2.9 – Crystals of Fbxo7^{F3}-Skp1

A. Diamond shaped crystals typical of the morphology of Fbxo7^{F3}-Skp1 crystals grown under vapour diffusion in 6% PEG 20K, 0.1M MES pH 6.5 and 0.005% Zwittergen 3.12. B. SDS-PAGE of protein used to grow crystals shows that this protein complex does not have a 1:1 stoichiometry when stained with Coomassie brilliant blue. Crystal photograph provided by Phillip Knowles.

2.4 Data collection from Fbxo7^{F3}-Skp1 crystals

The protein complex crystals described above are reproducible but they displayed anisotropic diffraction to only 8 - 12Å using a rotating anode X-ray generator as the X-ray source. This level of diffraction was not high enough to extract any information about these crystals.

Paratone was used as the main cryo-protectant for Fbxo7^{F3}-Skp1 crystals and produced a good glassy freeze with no ice rings visible. Cryo-protectant screens were also carried out using various concentrations of glycerol, PEG 400 and ethylene glycol to attempt to improve the diffraction of the crystals. Flash cooling in liquid nitrogen was also attempted in order to minimise the chance that the crystals were becoming disordered during the freezing process. None of these changes in cryo-protectant changed the diffraction quality of the crystals.

One explanation for the poor diffraction quality of these crystals is that the protein used to produce the crystals contained a number of high molecular weight contaminants (Figure 2.9b) at a low concentration, which could have been poisoning the crystal lattice. It is also clear from the SDS-PAGE that either Skp1ΔH8 does not stain well with Coomassie brilliant blue or that Fbxo7^{F3} and Skp1ΔH8 are not being produced in a 1:1 complex under these expression and purification procedures. If the latter is true then it is possible that the protein used for crystallisation is not mono-disperse and homogeneous and this would explain the difficulty in achieving crystals that diffract well.

In order to test the likelihood of the sample containing non-homogeneous material it was analysed by size exclusion chromatography as described in appendix A. All concentrated Fbxo7^{F3}-Skp1 samples analysed produced two peaks, one representing high molecular weight aggregates and the second a complex of 43-45kDa (Figure 2.10a and b). The anticipated molecular weight of a 1:1 complex of Fbxo7^{F3}-Skp1ΔH8 is 43kDa. SDS-PAGE (Figure 2.10c) shows that the protein eluted in the high molecular weight fractions consists of Fbxo7 and high molecular weight contaminants, which are likely to be *E. coli* heat shock proteins. The elution profile shown in Figure 2.10 is typical of all Fbxo7-Skp1 samples produced and analysed in this way. Figure 2.10a and b demonstrate that the extent of aggregation of Fbxo7 varies in different protein preparations however sample b is the most representative of the majority of samples.

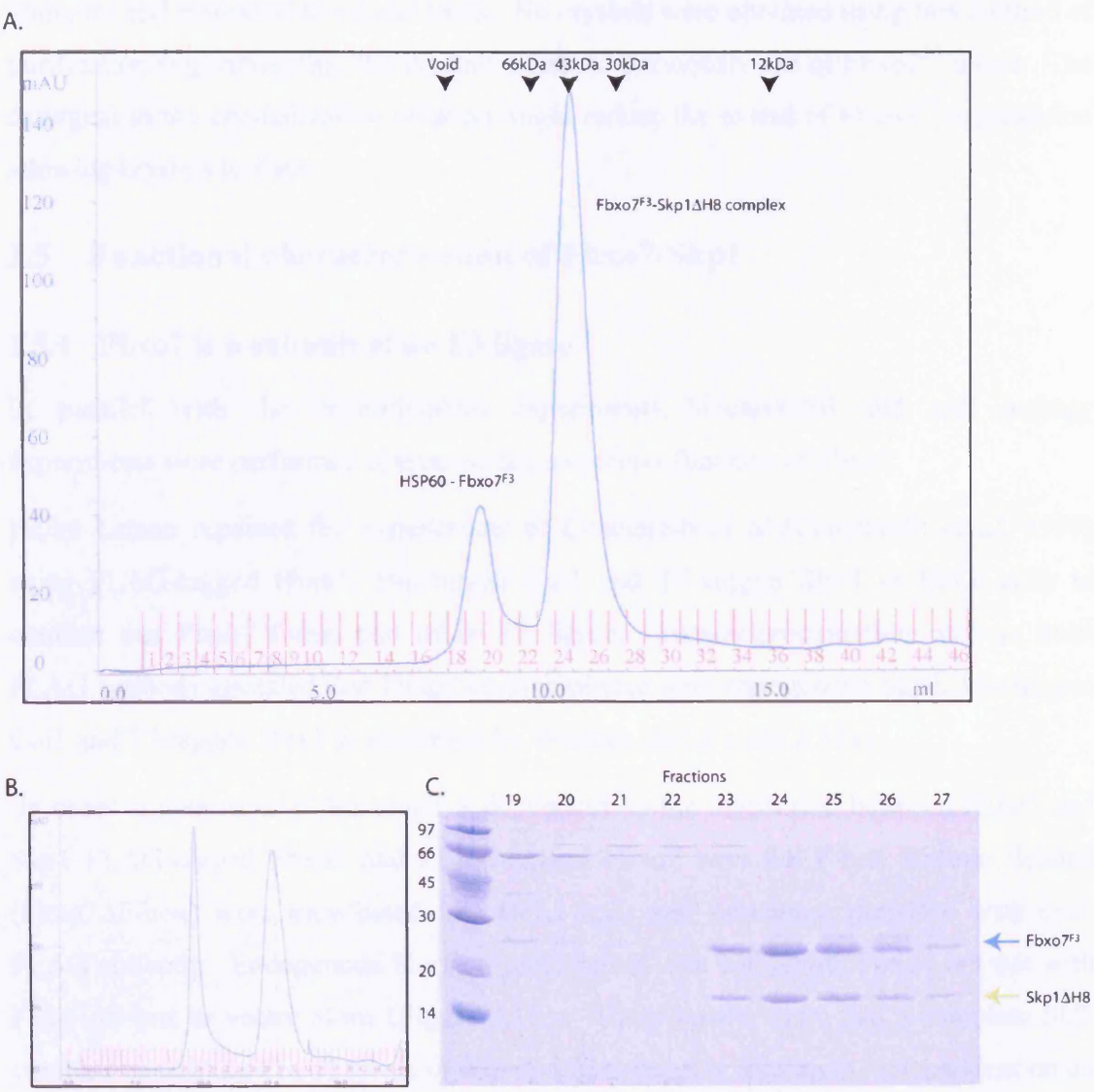

Figure 2.10 – Gel filtration profile of Fbxo7^{F3}-Skp1 used in crystallisation

Gel filtration performed using a Superdex 75 HR 10/30 column (Amersham Biosciences) at a rate of 0.4 ml/min using 20mM Tris-HCl pH 8.0 and 50mM NaCl as buffer. 500µl of protein sample was loaded onto the column. A) Gel filtration profile of Fbxo7^{F3}-Skp1 used in crystallisation showing a large molecular weight aggregate, likely to be Fbxo7^{F3} bound to an *E. coli* heat shock protein (identified as HSP60 by mass spectrometry and shown in Figure 2.12). B) Typical gel filtration profile, the ratios of the aggregate to complex are more typical here. C) SDS-PAGE of fractions from (A). This shows a high molecular weight contaminant in the aggregated fractions and a 1:1 ratio of Fbxo7-Skp1 in the complex fraction.

Gel filtered samples of the 43kDa Fbxo7^{F3}-Skp1 complex were concentrated to 16mg/ml and subjected to crystal trials. No crystals were obtained using this method of purification suggesting that the crystals obtained previously are of Fbxo7^{F3} alone. The detergent in the crystallisation solution might reduce the extent of Fbxo7^{F3} aggregation allowing crystals to form.

2.5 Functional characterisation of Fbxo7-Skp1

2.5.1 Fbxo7 is a subunit of an E3 ligase

In parallel with the crystallisation experiments biochemical and cell biology experiments were performed to analyse the molecular function of Fbxo7.

Heike Laman repeated the experiments of Cenciarelli *et al* (Cenciarelli *et al.* 1999) using FLAG-tagged Fbxo7, His-tagged Cul1 and T7-tagged Rbx1 in HeLa cells to confirm that Fbxo7 forms part of an E3 ligase. Immunoprecipitation with an anti-FLAG antibody revealed that Fbxo7 co-precipitated with endogenous Skp1, His-tagged Cul1 and T7-tagged Rbx1 as identified by Western blot (Figure 2.11a).

In order to determine if this result is dependent on the interaction between Fbxo7 and Skp1 FLAG-tagged Fbxo7 and FLAG-tagged Fbxo7 with the F-box domain deleted (Fbxo7ΔF-box) were transfected into HeLa cells and immunoprecipitated with anti-FLAG antibody. Endogenous Skp1 co-precipitated with full length Fbxo7 but not with Fbxo7ΔF-box or vector alone (Figure 2.11a). These results show that a complete SCF complex containing Fbxo7 forms *in vivo* and that complex formation is dependent on an intact F-box domain.

The interaction between Fbxo7 and the other components of the SCF has not previously been shown *in vitro*. In order to show this interaction *in vitro* Cul1-NTD (14-312) was applied to a glutathione Sepharose column charged with GST-Fbxo7^{F2}-Skp1 (data not shown) or GST-Fbxo7^{F3}-Skp1. The column was then fully washed and the resulting complex cleaved from the resin with thrombin. All proteins were expressed in *E. coli* and purified using standard protocols. In both constructs examined Cul1-NTD forms a complex with Fbxo7-Skp1 in the supernatant (Figure 2.11b). There is no difference between the constructs as would be expected as Cul1-NTD binds to Skp1 (Zheng *et al.* 2002) which is invariant in the two constructs.

The resulting Fbxo7-Skp1-Cul1 complexes were both submitted to crystal trials as described for the Fbxo7-Skp1 complex above. There were no crystals obtained from these screens. This could be because the samples were not applied to the additional gel filtration purification step that has been shown to be required for the Fbxo7-Skp1 complex alone to be homogeneous.

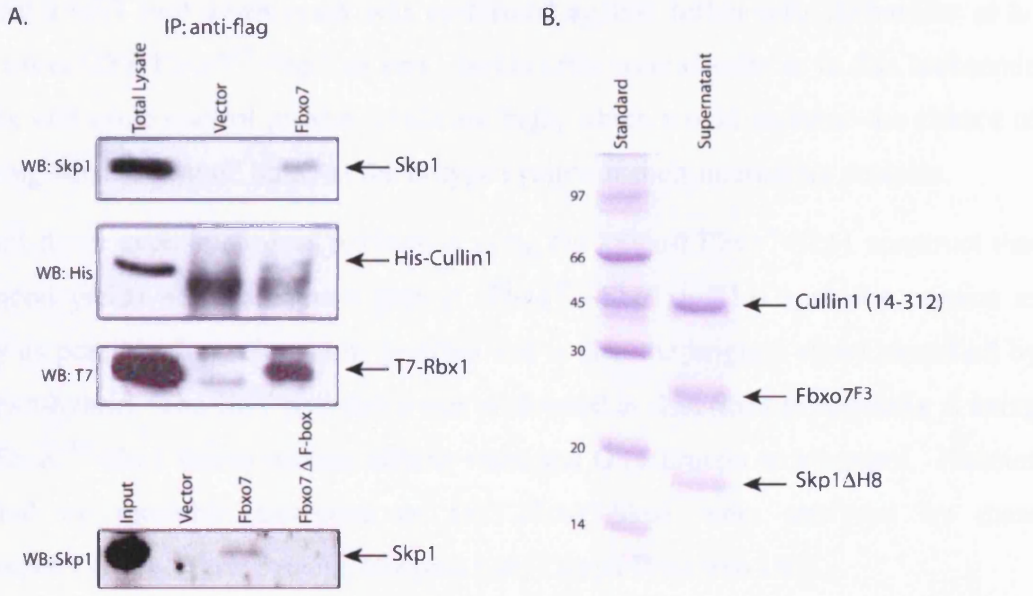

Figure 2.11 – Fbxo7 forms a complex with Skp1, Cul1 and Rbx1

A. Coimmunoprecipitation of FLAG-tagged Fbxo7 with endogenous Skp1, His-tagged Cul1 and T7-tagged Rbx1. When the F-box is deleted from Fbxo7 it no longer co-precipitates with Skp1. Cul1 contains a large non-specific band lower than His-Cul1 which is present in both vector alone control and Fbxo7 lanes. T7-Rbx1 also shows some non-specific binding of the antibody. B. Cul1-NTD binds to GST-Fbxo7^{F3}-Skp1 when immobilised on a glutathione Sepharose column. These three proteins co-elute in the supernatant, as seen by SDS-PAGE, when submitted to thrombin cleavage of the GST moiety. This shows that they form a ternary complex *in vitro*. Coimmunoprecipitation performed by Heike Laman.

2.5.2 Identification of Fbxo7-Skp1 binding partners from Jurkat cell lysates

One of the main challenges facing the study of E3 ubiquitin ligases is the identification of substrate proteins. At the start of this project no substrate or binding partner for Fbxo7 had been reported in the literature other than Skp1. The identification of proteins that interact with Fbxo7 would be an important step in the functional analysis of the SCF^{Fbxo7} complex.

Our collaborator Heike Laman identified that Fbxo7 interacts with a viral D-type cyclin by yeast two-hybrid interaction. When the cyclin encoded by *Herpesvirus samurai (HVS)* was used as bait a clone was isolated encoding the amino acids 129-393 of Fbxo7, a region containing both the F-box domain and the undefined globular domain. This interaction had not been confirmed as a direct interaction by *in vitro* analysis therefore a GST pull down assay was performed against Jurkat cells (Schneider *et al.* 1977) using GST-Fbxo7^{F2}-Skp1 as bait. Jurkat cells were chosen as in this leukaemia cell line cell cycle control protein levels are high, which would increase the chance of observing whether Fbxo7 binds to the D-type cyclins or their interacting proteins.

The pull-down experiment was performed using the longest Fbxo7-Skp1 construct that gave good yields of recombinant protein (Fbxo7^{F2}-Skp1). This complex mimics as closely as possible the endogenous proteins and is also the original clone identified by yeast two-hybrid. The GST pull down was performed as described in appendix A using GST-Fbxo7^{F2}-Skp1 fusion protein affinity resin and GST protein as a control. Proteins co-eluted by thrombin treatment of GST-Fbxo7-Skp1 were analysed by mass spectrometry (Nick Totty, Protein Analysis Lab, Cancer Research UK).

Fbxo7^{F2}-Skp1 interacts directly with endogenous Skp1 and Cullin1 (Figure 2.12a) as would be expected. This confirmation of a known interaction indicated that the protein binding partner identified by this approach might be physiologically relevant. Fbxo7^{F2}-Skp1 also binds to the proteasome inhibitor PI31 (and a truncated form of this protein), actin and *E. coli* HSP60. Actin is also bound in the GST control so this is likely to be an artefact of the experiment and *E. coli* HSP60 is presumably bound to Fbxo7-Skp1 from the original protein production due to the aggregated Fbxo7. PI31 has not been shown to interact with Fbxo7 or Skp1 before and so this represents a novel binding partner for Fbxo7-Skp1.

The GST pull-down did not reveal any evidence of interaction with the cell cycle control proteins cyclins or cyclin dependent kinases, as might have been expected from the yeast two-hybrid results. There is also no binding of the recently reported substrate of Fbxo7, HURP (Hsu *et al.* 2004). The interaction between HURP and Fbxo7 is thought to be via the C-terminal PRR of Fbxo7 and as this domain is missing from the recombinant protein the interaction is not detected.

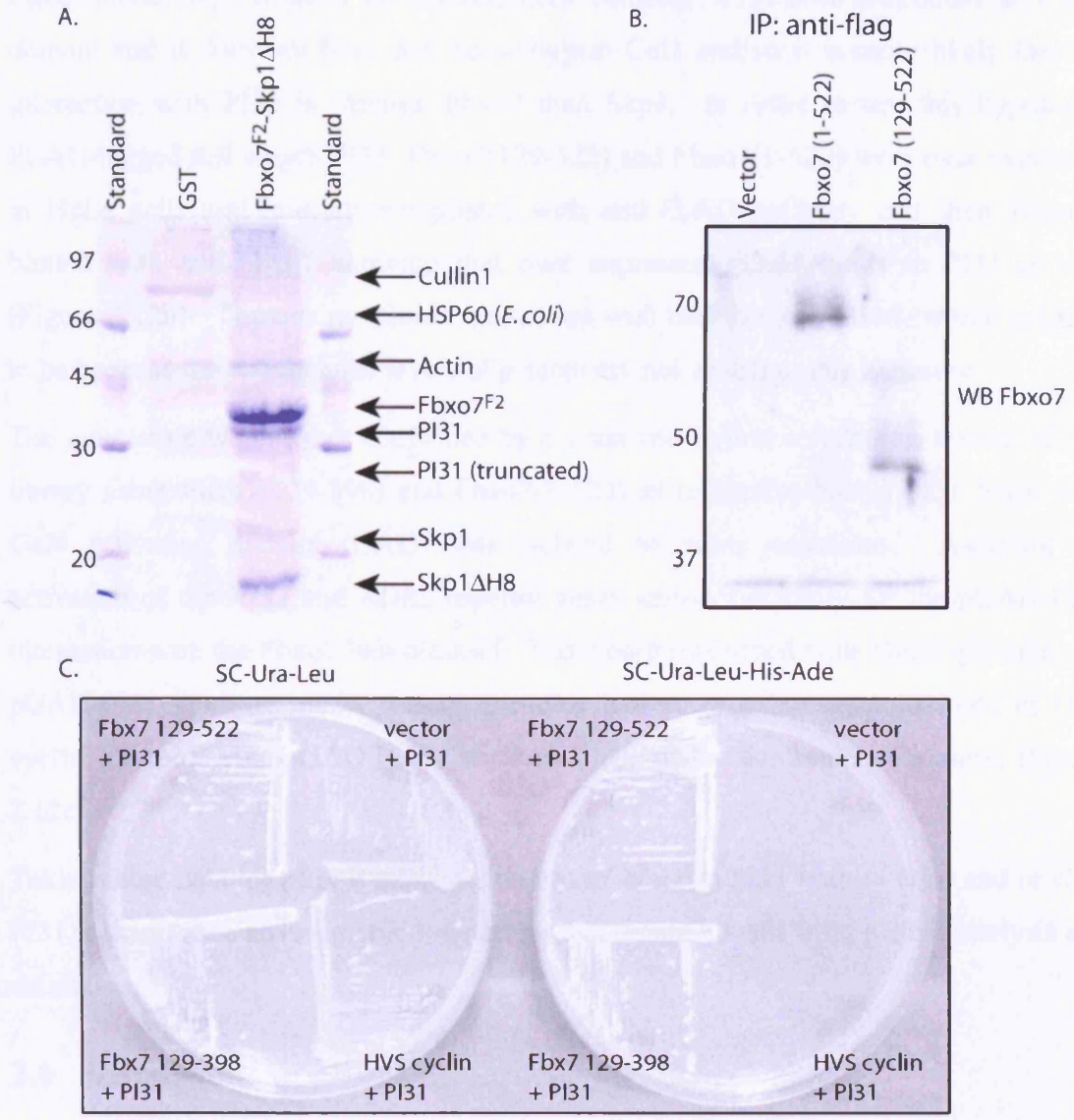

Figure 2.12 – Identification of PI31 as a binding partner of Fbxo7.

A. Identification of proteins from Jurkat cell lysates that co-elute with Fbxo7^{F2}-Skp1 upon thrombin cleavage of the GST-Fbxo7^{F2}-Skp1 fusion protein complex. Proteins were identified by SDS-PAGE and mass spectrometry. B. Fbxo7 co-immunoprecipitates with PI31. Fbxo7 was present in the anti-FLAG immunoprecipitates of PI31 as seen by Western blotting using anti-Fbxo7. C. pGAD-PI31 interacts specifically with Fbxo7 bait as shown by growth on media selecting for the plasmids (SC-Ura-Leu) and on media selecting for activation of reporter genes (SC-Ura-Leu-His-Ade). Coimmunoprecipitation and yeast-2 hybrid performed by Heike Laman.

The interaction between Fbxo7^{F2}-Skp1 and PI31 could be mediated through either Fbxo7 alone, Skp1 alone or the heterodimeric complex. PI31 does not contain an F-box domain and it does not have any homology to Cul1 and so it is more likely that the interaction with PI31 is through Fbxo7 than Skp1. In order to test this hypothesis FLAG-tagged full length PI31, Fbxo7(129-522) and Fbxo7(1-522) were over expressed in HeLa cells and immunoprecipitated with anti-FLAG antibody and then Western blotted with anti-Fbxo7 showing that over expressed Fbxo7 binds to PI31 *in vivo* (Figure 2.12b). There is no visible interaction with endogenous Fbxo7, which is likely to be because the endogenous levels of protein are not visible at this exposure.

The interaction was further confirmed by a yeast two-hybrid screen of a human cDNA library using Fbxo7(129-398) and Fbxo7(1-522) as respective baits. PI31 fused with Gal4 activation domain (GAD) was isolated on many occasions. Assaying for activation of the *HIS3* and *ADE2* reporter genes tested specificity for the pGAD-PI31 interaction with the Fbxo7 bait plasmid. Yeast co-transformed with Fbxo7 plasmid and pGAD-PI31 grew on media lacking histidine and adenine but plasmid alone or *HVS* cyclin plasmid with pGAD-PI31 failed to grow under the same conditions (Figure 2.12c).

Taking these data together it appears that Fbxo7 binds to PI31 both *in vitro* and *in vivo*. PI31 is therefore a novel interaction partner that would benefit from further analysis and validation.

2.6 Discussion

The goals for the use of the recombinant proteins described in this chapter were to identify novel binding partners of Fbxo7 and to crystallise and solve the structure of the Fbxo7-Skp1 complex. The purified recombinant proteins Fbxo7-Skp1 were produced at moderate levels allowing crystallisation trials and biochemical analysis of the complex. These proteins will be useful tools in the analysis of the novel interaction between Fbxo7-Skp1 and PI31.

The identification of aggregated Fbxo7 in the Fbxo7-Skp1 protein complex used to obtain Fbxo7-Skp1 crystals indicates that this protein was not homogeneous. The possibility that Fbxo7 may crystallise in the absence of Skp1 should be considered and this may aid the crystallisation of an Fbxo7-substrate complex.

The observation that Fbxo7 binds to the Skp1 containing the loop deletions used for previous structure determination of Skp1-F-box proteins (section 2.2.1) is important since the interaction between Skp1 and other F-box proteins is not identical in the three previously solved structures. Skp2, a LRR F-box protein (Schulman *et al.* 2000), and β-TRCP, a WD40 F-box protein (Wu *et al.* 2003), bind to Skp1 via a four helix bundle with the F-box domain comprising a loop region (L1) and 3 α helices (α1 – α3). However cdc4, another WD40 F-box protein (Orlicky *et al.* 2003), has a different F-box topology comprising five α helices (α0 – α4) and although it interacts with Skp1 using a similar four helix bundle the Skp1 α helix 8 and F-box α4 deviate significantly. The fact that this loop region, when deleted in Skp1, has no effect on the binding of Skp2, β-TRCP, cdc4 and Fbxo7 indicates that although the topology of the F-box itself may vary this deletion does not have any effect on Skp1-F-box binding irrespective of which class of F-box protein is involved. The loop deletion construct with reasonable expression (Fbxo7-Skp1^{F10}) may be worth re-visiting now that technology has improved and the volume of high concentration protein required for screening 96 crystallisation conditions using the mosquito™ robot (TPP Labtech) is only 12μl.

The identification of the proline rich region at the C-terminus of Fbxo7 and its reported role as a protein-protein interaction domain (Hsu *et al.* 2004) indicates that this region may be important in the function of Fbxo7. Further work on the analysis of the role of this region *in vitro* with respect to the interaction with PI31 may be valuable. It might be useful to identify the minimal region of Fbxo7 that interacts with PI31 and to evaluate the role that the PRR may play in this interaction.

The Ubl domain of Fbxo7 may also play a role in the function of this protein *in vivo* because Ubl domains have previously been shown to interact directly with the ubiquitin-interacting motif (UIM) of S5a, a subunit of the 26S proteasome. The structure of the UIM of S5a bound to the Ubl of HR23B showed that the Ubl adopts the ubiquitin fold and that the UIM contact sites in the Ubl are conserved in ubiquitin (Fujiwara *et al.* 2004). This indicated that the Ubl and ubiquitin make the same contacts and so Ubls can mimic ubiquitin in its interactions. It is interesting that Fbxo7, a protein involved in the ubiquitination pathway, also contains this motif. In addition to this the C-terminal region of the Ubl domain in Fbxo7 is conserved and so it may be cleaved at this site and could provide a different regulatory mode. One of the plant

Fbxo7 proteins does not contain the Ubl domain and so this form of the protein could represent a function for the Fbxo7 protein in the absence of the Ubl.

The identification of Fbxo7 as a subunit of an E3 ligase both *in vitro* and *in vivo* reveals that this complex can form in the absence of any co-factors and so is likely to be an active E3 ligase although such an activity remains to be reconstituted *in vitro*. A surprising finding was the identification of a novel binding partner for Fbxo7 but not the reported interacting partner Cdk6. Other known interacting partners such as Cullin1 were identified and so it is likely that this assay is able to detect physiological interactions. It is possible that the full length Fbxo7 is required for interaction with Cdk6. The identification of PI31 as a novel binding partner could indicate several roles for this interaction. It is possible that PI31 could be a substrate of SCF^{Fbxo7} or it could be a co-factor, binding to Fbxo7 in order to bring another protein into position for polyubiquitination and degradation.

The interaction between PI31 and Fbxo7 will be discussed later in this thesis.

This chapter shows that Fbxo7 can form a complete SCF both *in vitro* and *in vivo*. It was also found that the proteasome inhibitor PI31 acts as a binding partner to Fbxo7 both *in vitro* and *in vivo* and that the domain mediating this interaction is within amino acids 129-398 of Fbxo7.

3 Production and structure determination of the PI31 N-terminal domain

Following the identification of the structurally uncharacterised protein PI31 as a binding partner for Fbxo7 its suitability as a crystallisation target was assessed. Sequence analysis from 3D-PSSM and CD data (McCutchen-Maloney *et al.* 2000) was used to predict that PI31 has a globular domain spanning residues 1-151 which would be a suitable crystallisation target. This chapter describes the structure determination of this globular domain of PI31. A flowchart describing the experiments undertaken can be found in Figure 3.1.

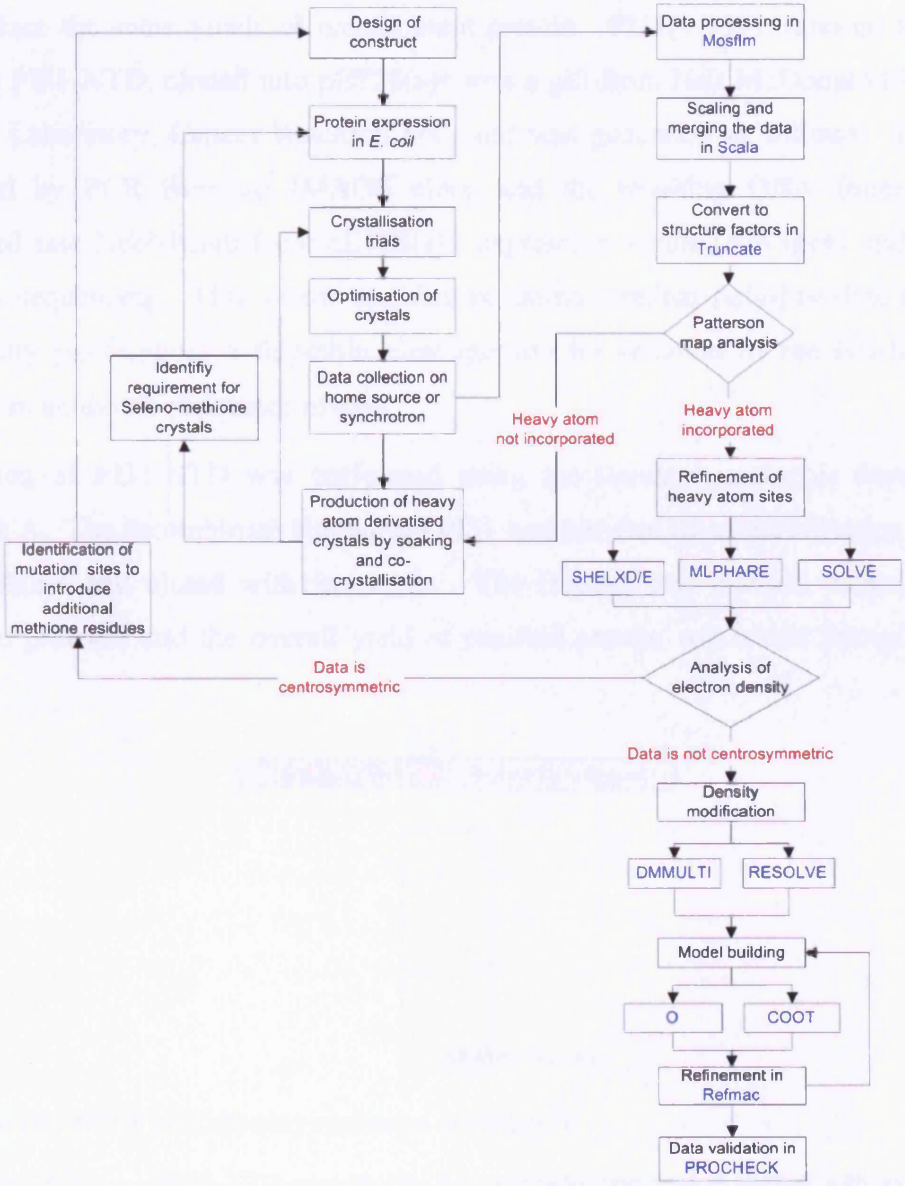

Figure 3.1 – Flowchart of the experiments described in chapter three.

3.1 Cloning and expression of PI31

Circular dichroism (CD) analysis of recombinant PI31 performed by McCutchen-Maloney *et al* (McCutchen-Maloney *et al.* 2000) identified two separate domains within the protein. PI31 full length protein displayed elements of both α helical and random-coil structure whereas PI31(1-192) showed predominately α helical structure and PI31(192-271) exhibited a random-coil structure. The C-terminal domain of PI31 consists of a PRR in which 26% of the terminal 117 amino acids are proline residues, which is reminiscent of the C-terminus of Fbxo7. McCutchen-Maloney *et al* (McCutchen-Maloney *et al.* 2000) previously reported that whereas high expression of PI31 protein in *E. coli* was achieved using the pET expression system other systems did not produce the same yields of recombinant protein. PI31(1-151), referred to in this thesis as PI31-NTD, cloned into pET28(a)+ was a gift from Neil McDonald (Structural Biology Laboratory, Cancer Research UK) and was generated as follows: PI31 was amplified by PCR from an IMAGE clone and the resulting DNA fragment was subcloned into NdeI/HindIII cut pET28(a)+ expression vector (Novagen) and verified by DNA sequencing. This vector contains an amino-terminal poly-histidine (His) tag for affinity purification, a thrombin cleavage site for removal of the His-tag and a kanamycin antibiotic resistance marker.

Expression of PI31-NTD was performed using the standard protocols described in appendix A. The recombinant His-tagged PI31 was purified by immobilisation on a Ni-NTA column and eluted with imidazole. The His-tag was cleaved overnight with thrombin protease and the overall yield of purified protein was about 50mg/l (Figure 3.2).

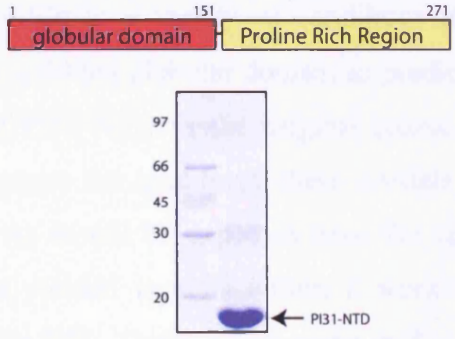

Figure 3.2 – SDS-PAGE showing expression of PI31-NTD

Expected position of PI31-NTD generated by the expression construct is marked with an arrow. This construct produced high yields of pure protein.

3.2 Crystallisation of PI31

Crystallisation trials were carried out using PI31-NTD recombinant protein at a concentration of 6-8mg/ml in 20mM Tris-HCl pH 8.0 and 5mM 2-ME as described in appendix B. In Hampton screens 1 and 2 and Wizard screen 1 several conditions yielded crystals as listed in Table 3.1.

Screen Name	Number	Description	Time	Contents of Crystallisation solution
Hampton One	10	Showers of crystals and protein precipitate	24 hours	30% w/v PEG 4000, 0.1M sodium acetate trihydrate pH 4.6, 0.2M ammonium acetate
	35	Crystalline needles	1 week	0.8M mono-sodium dihydrogen phosphate, 0.8M mono-potassium dihydrogen phosphate, 0.1M HEPES pH 7.5
	37	Showers of crystals	1 hour	8% w/v PEG 4000, 0.1M sodium acetate trihydrate pH 4.6
	42	Crystalline diamond plates	24 hours	20% w/v PEG 8000, 0.05M mono-potassium dihydrogen phosphate
Hampton Two	7	Showers of crystals	1 week	10% w/v PEG 1000, 10% w/v PEG 8000
Wizard One	17	Showers of crystals and protein precipitate	24 hours	30% w/v PEG 8000, 0.1M sodium acetate pH 4.5, 0.2M Li_2SO_4
	19	Showers of crystals	24 hours	20% w/v PEG 1000, 0.1M Tris-HCl pH 7.0,
	30	Showers of crystals	1 week	1.26M $(NH_4)_2SO_4$, sodium acetate pH 4.5, 0.2M NaCl
	31	Showers of crystals and protein precipitate	1 week	20% w/v PEG 8000, 0.1M HEPES pH 7.5, 0.2M NaCl

Table 3.1 –Initial crystallisation hits from sparse matrix screens using vapour diffusion

Screen Name	Number	Description	Time	Contents of Crystallisation solution
PEG/Ion	7	Crystalline needles	1 week	20% w/v PEG 3350, 0.2M calcium chloride dehydrate
	9	Crystalline diamond plates	72 hours	20% w/v PEG 3350, 0.2M ammonium chloride
	12	Crystalline diamond plates	1 week	20% w/v PEG 3350, 0.2M ammonium iodide
	16	Showers of crystals and protein precipitate	72 hours	20% w/v PEG 3350, 0.2M magnesium nitrate hexahydrate
	39	Showers of crystals	72 hours	20% w/v PEG 3350, 0.2M sodium dihydrogren phosphate monohydrate
	41	Showers of crystals	72 hours	20% w/v PEG 3350, 0.2 M potassium dihydrogen phosphate
	43	Showers of crystals	72 hours	20% w/v PEG 3350, 0.2 M ammonium dihydrogen phosphate

Table 3.2 – Crystallisation hits from Hampton PEG/Ion screen using vapour diffusion

PI31-NTD crystallised readily in a variety of conditions indicating that the domain boundaries used represent a folded globular domain as predicted. 7 of the 9 conditions that generated crystals of PI31-NTD in the original sparse matrix screens contained PEG and in order to improve the quality of these crystals the Hampton PEG/Ion™ screen was carried out. As would be expected from the results of the sparse-matrix screen several conditions yielded crystals within a week of setting up the vapour diffusion experiment (Table 3.2). Optimisation using grid screens designed around the conditions highlighted by the PEG/Ion screen produced diffraction-size crystals of PI31-NTD with a plate diamond morphology in 0.1M ammonium iodide (NH_4I) and 20-22% PEG 3350 (Figure 3.3).

Figure 3.3 – Crystals of PI31-NTD

Protein crystals of PI31-NTD grown in 0.1M NH_4I and 20% PEG 3350. The storage buffer for the protein contained 20mM Tris-HCl pH 8.0 and 5mM 2-ME. These crystals were reproducible for different preparations of crystal solution and for different protein expression batches.

3.3 Data collection and analysis of PI31-NTD crystals

Data collection was performed on several crystals of PI31-NTD using both rotating anode and synchrotron radiation sources. All PI31-NTD data collected had a high mosaic spread, which was dependent on the orientation of the crystal and so varied with a sinusoidal pattern during the rotation of the crystal. Figure 3.4 shows typical diffraction images of a PI31-NTD crystal taken on our home source. There are a considerable number of partial reflections visible as clearly weak spots at the edges of the lunes.

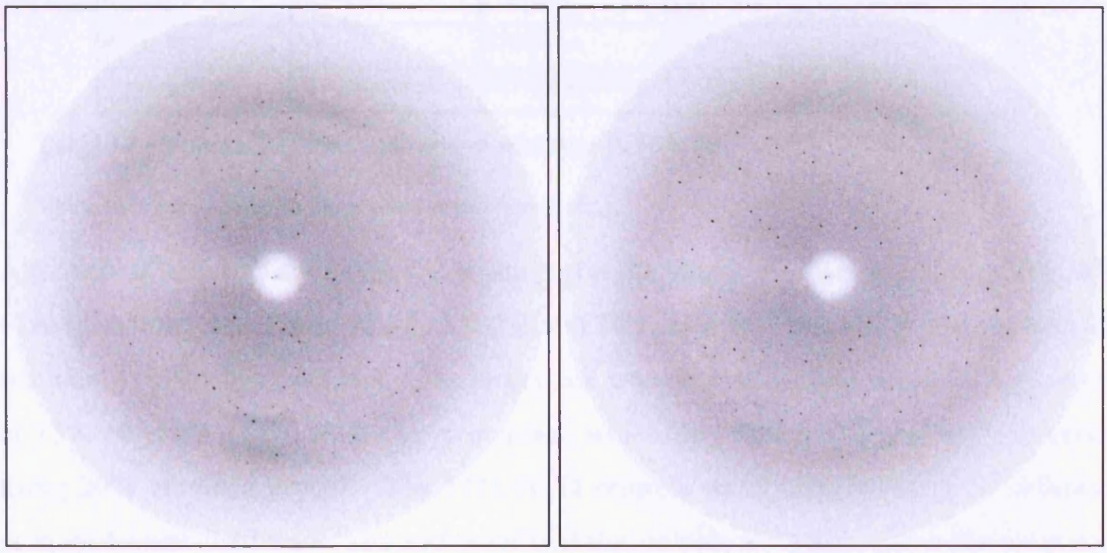

Figure 3.4 – Diffraction data images of PI31-NTD

Data from this crystal was collected on our home source. These images are taken at 0° and 90° and show the quality of the diffraction images obtained with a resolution of 2.6Å at the edge.

3.4 Processing and analysis of PI31-NTD diffraction data

All the PI31-NTD crystals had a monoclinic space group (Hahn 1996). There were two different unit cells identified during the data collection (Table 3.3). Crystal parameters calculated by the program MOSFLM (described in appendix B and (Leslie 1992)) were used to predict diffraction spots. Visual comparison of the observed and predicted spots was used to validate the calculated parameters. This revealed that the predicted mosaic spread calculated for PI31-NTD was too low and a new value was applied based on the best fit for the data (0.8° - 1.2°). An attempt was made to refine this new value but unfortunately due to the large variance in mosaicity (0.5° - 1.5°) this rendered the refinement unstable. The mosaic spread was refined for every batch of ten images and so varied throughout the integration of the images. The value chosen by best fit was a starting value, which led to a more stable integration than that generated automatically.

Dataset number	I	II	III	IV
Data collection site	Daresbury 14.2	Home source	ESRF BM29	ESRF 14.4
Wavelength (Å)	0.971	1.541	0.979	0.979
Space group	C2	C2	C2	C2
Unit cell	a=108.5 b=40.8 c=65.1 β=109.5	a=109.5 b=45.2 c=67.3 β=111.5	a=109.2 b=43.0 c=132.2 β=109.34	a=108.9 b=43.4 c=132.3 β=109.4
Resolution range (Å)	37.8-2.5 (2.64-2.5)	28.1-2.64 (2.78-2.64)	54.2-2.8 (2.95-2.8)	100-2.1 (2.21-2.1)
Completeness (%)	94.4	99.9	99.7	92.2
Rmeas (%)	9.7 (43.2)	8.5 (41.9)	4.5 (9.1)	5.4 (20.7)
<I>/sd	12.7 (2.9)	21.1 (4.8)	34.3 (18.5)	14.0 (3.9)
Multiplicity	3.6 (3.7)	7.2 (5.0)	7.1 (6.9)	1.8 (1.7)
Cryo-protectant	20% Ethylene Glycol	20% Ethylene Glycol	15% Ethylene Glycol	20% Glycerol

Table 3.3 - Summary of data collected of wild type PI31-NTD

Value in brackets refer to the highest resolution shell

Although all datasets are in the C2 space group the length of the c axis is variable, 65-67Å (datasets I and II) or 132-133Å (III and IV), with the remaining cell parameters remaining reasonably constant. The larger cell was observed when using 20% glycerol or 15% ethylene glycol as a cryo-protectant, while the smaller cell was only observed using 20% ethylene glycol. The PI31-NTD crystals were very sensitive to different cryo-protectant conditions, and additionally if the crystals were exposed to the air at any time during the mounting process all diffraction disappeared. Many different cryo-protectant conditions were tested in an attempt to minimise the mosaicity problem, but only glycerol and ethylene glycol could be used as a cryo-protectant to obtain

diffraction to better than 3Å. Analysis of the Matthew's co-efficient ((Matthews 1968) and appendix B) revealed that the smaller unit cell contained 2 molecules per AU and the larger unit cell contained 4 molecules per AU.

Native Patterson map analysis of the large cell data shows a clear peak at u=0,v=0,w=0.5 indicating translational non-crystallographic symmetry (NCS). The small and large unit cell crystal forms may be dependent on whether the symmetry is NCS (large cell) or crystallographic (small cell).

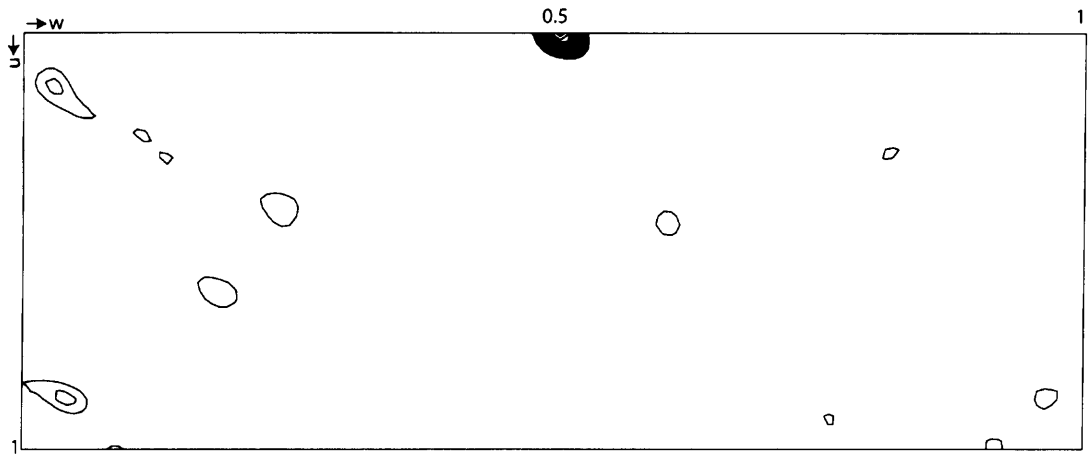

Figure 3.5 – Native Patterson map on the Harker section (v=0) for large c cell axis.

This Patterson map includes data from 20 – 6Å, it is contoured at 2σ with an interval of 1σ. The large peak at u=0, v=0, w=0.5 indicates the NCS along this axis.

The completeness as calculated by SCALA (Evans 1993, 1997) is adequate for all the datasets collected (varying from 99.9 - 92.2%) and the multiplicity (measure for the number of times each reflection is measured) is high for all datasets except for that dataset collected at ESRF 14.4 where the multiplicity is only 1.8. <I>/sd is a measure of the signal to noise ratio and it is above the recommended value of 2 for all of the datasets with the lowest value being 2.9 in the highest resolution shell of dataset I.

Of the two datasets (I and II) collected containing 2 molecules per AU, the data collected at the home source (II) has a similar resolution to that collected at Daresbury (I). However the remaining statistics are better for dataset II with a completeness of 99.9%, a multiplicity of 7.2 and a higher <I>/sd in the highest resolution shell, indicating that the data is stronger and more redundant and therefore a good dataset for refinement. The crystallisation condition for this crystal contained a peptide from the C-terminus of PI31, which is not visible in the electron density. Subsequent biochemical work showed that this peptide did not bind to PI31-NTD (data not shown).

Due to the superior quality of the data all data analysis of native datasets was performed on dataset II.

The program TRUNCATE (French and Wilson 1978) was used to calculate the structure factor amplitudes from the average intensities from the PI31-NTD crystals. The corresponding Wilson plot is shown in Figure 3.6 and the values calculated for the scale factor (C) and the temperature factor (B) were C = 11.0 and B = 60.7Å^2. The structure factor amplitudes calculated in this process were used in all following crystallographic analysis.

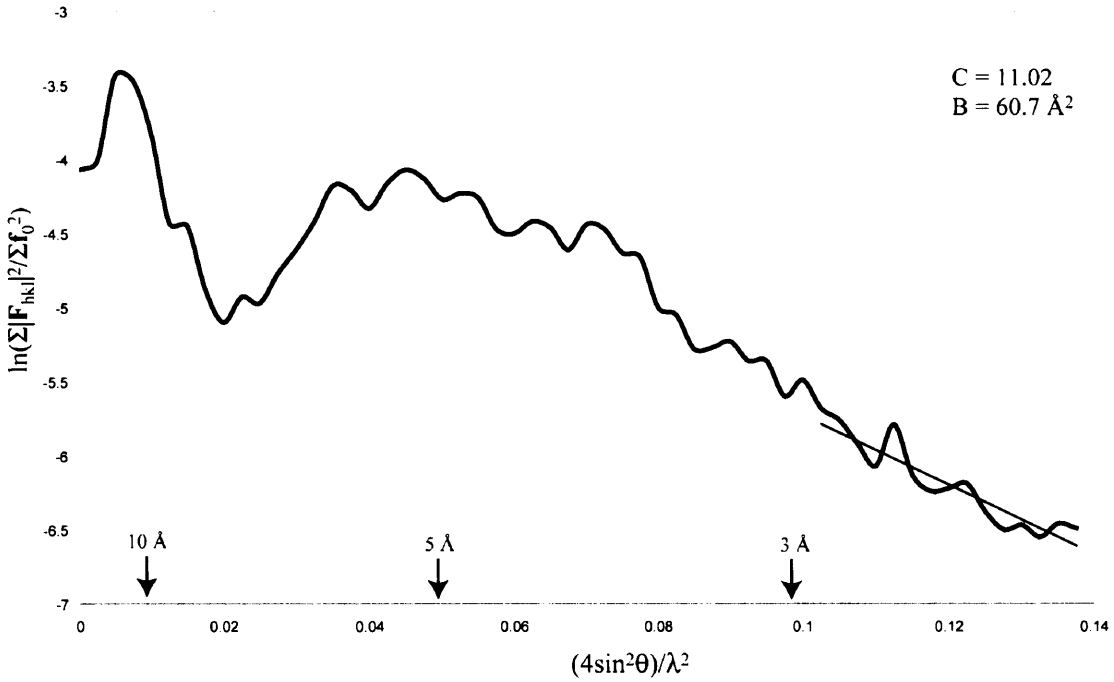

Figure 3.6 - Wilson plot of PI31-NTD diffraction data (dataset II)

Wilson plot calculated for dataset II using TRUNCATE (Collaborative Computational Project 1994). A straight line was fitted to the plot above 3Å where the linear relationship is valid and the intercept and gradient of this used to estimate the scale factor and temperature factor.

3.5 Phase determination

3.5.1 Heavy atom soaks

Having obtained reasonable resolution native data the phase problem needed to be addressed. The first method attempted was SIR but since PI31-NTD crystallised in NH_4I a large number of the frequently used heavy atoms precipitated in the crystallisation solution. This is partly because many of the heavy atom compounds used

reacted with the NH_4I to produce insoluble metal compounds. All heavy atom compounds that were soluble in the crystallisation solution were tested on PI31-NTD crystals. Three compounds which did not cause visible destruction to the crystals were ethylmercurithiosalicylic acid (EMTS), mersalyl acid and potassium tetracyanoplatinate (II). Crystals soaked in each of these compounds were tested for diffraction at SRS Daresbury beamline 14.1 using different concentrations of heavy atom compound and different lengths of soak. This beamline allowed the wavelength to be tuned to obtain the highest possible anomalous signal from the heavy atoms. The only crystals that still diffracted x-rays were those soaked in EMTS.

The data collected at a wavelength of 1Å after soaking crystals at a concentration of either 1mM or 5mM EMTS for periods of 1 hour up to 48 hours maintained the same space group and unit cell as the native crystal. Unfortunately on analysis of the Patterson maps no peaks due to the heavy atoms could be found and so it was unlikely that any well ordered Hg atoms existed in the crystal.

After discussion it was thought that the new method of using halides as a phasing vehicle could be used in this case as the crystallisation solution contained iodide. It has been shown by Dauter *et al* (Dauter *et al.* 2000) that protein crystals soaked for a short time in a cryo-solution containing iodide ions can lead to the incorporation of this anomalous scatterer into the ordered solvent region and this can then be used in a SAD experiment. It is not possible to use iodide in a MAD experiment as synchrotrons cannot be tuned to the wavelengths that would be required for this approach because the K edge is at 0.37Å and L_I edge at 2.39Å, parameters that are beyond the capability of most beamlines (Carugo *et al.* 2005). Using X-rays with a wavelength of 1.54Å, the wavelength from a rotating anode, the anomalous signal for iodide is $f'' = 6.8$ electrons (Dauter and Dauter 2001) which is quite a significant signal and has been used for structure solutions. Unfortunately despite the fact that ammonium iodide significantly improves the quality of the PI31-NTD crystals it was not possible to detect any anomalous signal using anomalous difference Patterson maps that could be attributed to the iodide in the solution, so no iodide ions were sufficiently ordered to be used for phasing.

3.5.2 Selenomethionine phasing using SAD

Production of protein labelled with heavy atoms such as selenium and subsequent crystallisation of this modified protein introduces anomalously scattering atoms into the crystal. Methionine labelled with selenium is commercially available and by using supplemented minimal media a selenomethionine labelled PI31-NTD protein was produced (appendix A). PI31-NTD Se-Met protein was produced at a concentration of about 20mg/l in the form of a single species of the expected molecular weight when analysed by gel filtration (data not shown).

The crystals of PI31-NTD Se-Met formed in the same conditions as the native protein, with a slightly higher concentration of PEG 3350 producing the best crystals (22% as compared to 20% previously). The crystals have the same morphology (Figure 3.7b and c) as the wild type protein and the diffraction quality of the Se-Met crystals was comparable to the wild type protein at about 2.5Å (Table 3.4). An absorbance scan performed on the PI31-NTD Se-Met crystal revealed that the selenomethionine had incorporated as it produced the characteristic absorbance for selenium (Figure 3.7d). The peak energy was revealed to be 12667eV and this value was used to calculate the optimum wavelength for data collection.

Dataset number	V
Data collection site	Daresbury 14.1
Wavelength (Å)	0.978
Space group	C2
Unit cell	a = 109.0, b = 43.2, c = 132.6, β = 109.2
Resolution range (Å)	48.22 - 2.5 (2.64 - 2.5)
Completeness (%)	98.2
Rmeas (%)	6.5 (15.0)
Rano (%)	4.3 (7.3)
<I>/sd	20.4 (9.2)
Multiplicity	7 (6.7)
Cryo-protectant	20% Glycerol

Table 3.4 – Data processing statistics of Se-Met PI31-NTD crystals

Values in brackets refer to the highest resolution shell.

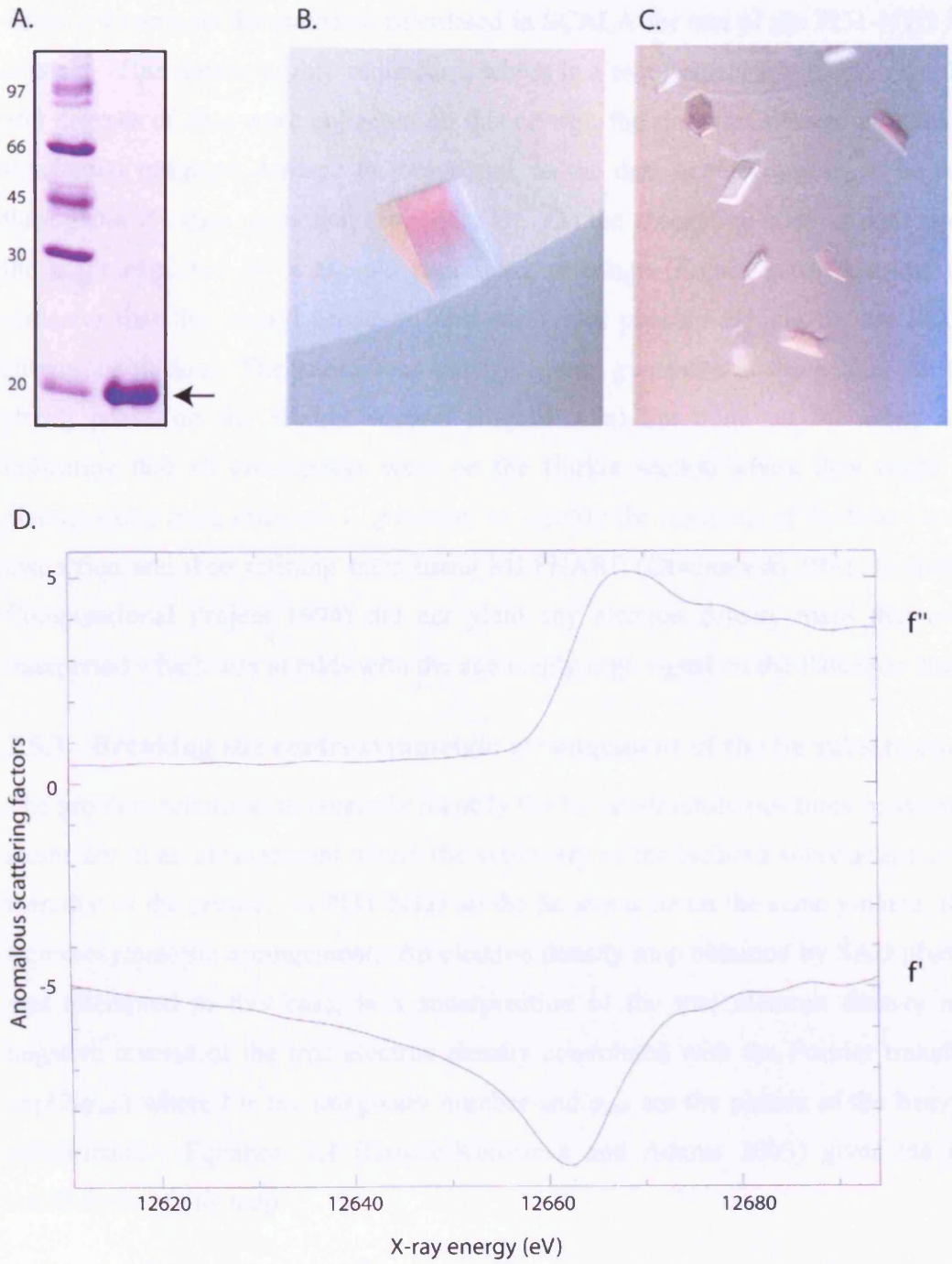

Figure 3.7 – Se-Met crystals of PI31-NTD

A) The black arrow marks the expected position of the His-tagged PI31-NTD Se-Met protein. B) Crystals formed in 0.1 M Ammonium iodide and 22% PEG 3350 and have a very similar morphology to the wild type protein crystals (C, as in Figure 3.3). D) Absorbance scan of PI31-NTD Se-Met crystal showing the characteristic selenium absorbance profile.

Table 3.4 contains the statistics calculated in SCALA for one of the PI31-NTD Se-Met crystals. This data is highly redundant, which is a requirement for anomalous phasing. 360 degrees of data were collected on this crystal; the statistics showed no evidence of substantial radiation damage to the crystal, as the data quality appears to be constant throughout the data collection (appendix B). On the absorption scan an edge is seen at the point expected for a sample containing selenium (Figure 3.7d) and this was an indicator that this would prove to be a successful phasing vehicle for the PI31-NTD structure solution. The anomalous Patterson map generated at the beam line showed strong peaks on the Harker section (Figure 3.9a) but none on the other sections indicating that all cross-peaks were on the Harker section where they could not be distinguished from self-peaks. Attempts to identify the positions of the heavy atoms by inspection and then refining them using MLPHARE (Otwinowski 1991; Collaborative Computational Project 1994) did not yield any electron density maps that could be interpreted which was at odds with the seemingly high signal on the Patterson map.

3.5.3 Breaking the centrosymmetric arrangement of the Se substructure

The problem remained to correctly identify the Se substructure positions because the Se atoms are in an arrangement where the symmetry of the isolated substructure is higher than that of the crystal. In PI31-NTD all the Se atoms lie on the same y-plane, forming a centrosymmetric arrangement. An electron density map obtained by SAD phasing, as was attempted in this case, is a superposition of the true electron density and the negative inverse of the true electron density convoluted with the Fourier transform of $\exp(2i\varphi_{sub})$ where i is the imaginary number and φ_{sub} are the phases of the heavy atom substructure. Equation 3.1 (Grosse-Kunstleve and Adams 2003) gives the Fourier coefficients of this map.

Equation 3.1 $$F - F^* \exp(2i\varphi_{sub})$$

The second term of Equation 3.1 is expected to contribute noise to the background in the SAD map that can be identified and removed by density modification and solvent flattening. The case of a centrosymmetric substructure in a non-centrosymmetric space group is different. If the centre for inversion of the substructure is placed at the origin of the unit cell all phases φ_{sub} are either 0° or 180° and therefore $\exp(2i\varphi_{sub}) = 1$. This reduces Equation 3.1 to $F - F^*$ which means that the SAD map will now be the

superposition of the true electron density with its exact negative inverse and interpretation of the map will be very problematic (Grosse-Kunstleve and Adams 2003).

Several strategies were identified to overcome this problem. One approach would be to use MAD phasing, in which case the ambiguity in the phases is removed and the true phases of the substructure can be found. However this is still a more complex procedure than solving a MAD structure without centrosymmetry and so a way to break the centrosymmetric arrangement and increase the anomalous signal was sought. It has been shown that the addition of a small number of methionines can simplify the location of Se sites for use in MAD phasing (Gassner and Matthews 1999). The optimal amino acid for substitution is leucine, followed by isoleucine; neither of these mutations had a significant effect on the structure of lysozyme (Gassner and Matthews 1999; Ohmura *et al.* 2001).

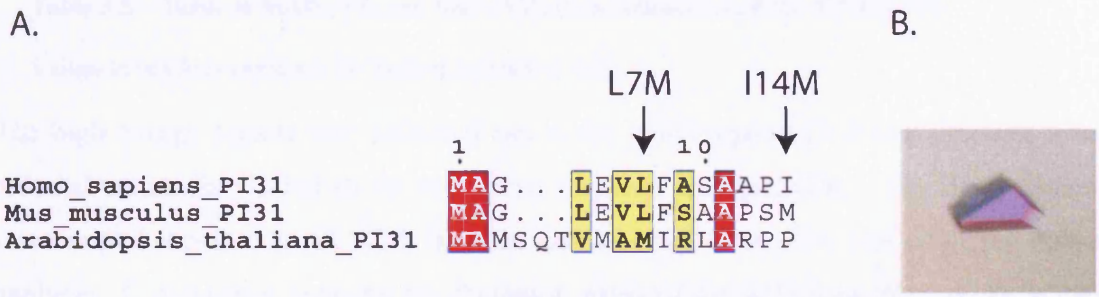

Figure 3.8 – N-terminal PI31 alignment and crystal of PI31-NTD (L7M)

A. Alignment of amino terminal 14 residues of PI31 with the mutations indicated. Alignment produced using ESPript (Gouet *et al.* 1999) B. Photograph of a PI31-NTD L7M crystal

Sequence comparison of PI31 using the program ClustalX (Thompson *et al.* 1997) showed that the conservation and alignment of both Leu7 and Ile14 with methionines was maintained in other species including plants (Figure 3.8) and so represented possible mutation sites that were likely to maintain the structural integrity of the molecule. Two mutant constructs (PI31-NTD (L7M) and PI31-NTD (I14M)) were generated, using the Quikchange protocol outlined in appendix A, and sequence confirmed by DNA sequencing. Protein with Se-Met incorporated was produced for both mutants with a yield comparable to that of wild type Se-Met protein. The mutant proteins were of the expected molecular weight when analysed by gel filtration (data not shown). Both proteins were submitted to crystallisation trials but only PI31-NTD (L7M) produced crystals (Figure 3.8b) in the same conditions as for the wild type

protein. Two Se-Met incorporated PI31-NTD (L7M) crystals were used to collect datasets at BM14 at the ESRF and the statistics for these are outlined in Table 3.5.

Dataset number	**VI**			**VII**
Crystal	**One**			**Two**
Space group	C2			C2
Unit cell	a = 108.4 , b = 42.8, c = 66.6, β = 109.1			a = 108.7, b = 42.3, c = 66.3, β = 109.4
Resolution range (Å)	39.75 - 2.5 (2.64 - 2.5)			27.12 - 2.50 (2.64 - 2.5)
Cryo-protectant	20% Ethylene Glycol			20% Ethylene Glycol
Wavelength (Å)	0.97855 (peak)	0.9790 (inflection)	0.9770 (high energy)	0.97855 (peak)
Rmeas (%)	7.2 (16.9)	6.1 (19.3)	7.1 (18.6)	4.7 (15.1)
Rano (%)	4.5 (8.0)	3.9 (10)	4.8 (11.1)	4.7 (8.4)
<I>/sd	25.5 (12.5)	18.1 (8.6)	16.2 (7.3)	31.9 (11.8)
Multiplicity	7.3 (7.4)	3.7 (3.7)	3.7 (3.7)	7.2 (7.3)
Completeness (%)	98.4	98.1	98.2	99.9

Table 3.5 – Table of MAD (VI) and SAD (VII) data collected on PI31-NTD (L7M)

Values in brackets represent the highest resolution shell.

The high energy remote was collected last in the MAD dataset (VI) and although it has reduced values for <I>/sd all the remaining statistics are acceptable. The three datasets were scaled together in SCALA and the output after truncation was used for further analysis. Comparison between the Patterson maps of the wild type PI31-NTD Se-Met protein and the L7M PI31-NTD protein showed that there were new self- and cross-peaks on the Harker section (Figure 3.9c and d) and cross-peaks were visible on other sections showing that the mutated residue was ordered and the Se position could be used in the determination of the phases for the heavy atom substructure (Figure 3.9b).

There are three methionine residues per mutant PI31-NTD molecule. Both the crystals from which data were collected were in the smaller space group with 2 molecules per AU and therefore there were six Se atoms to locate. SHELXD was used by Judith Murray-Rust (Structural Biology Laboratory, Cancer Research UK) to calculate the positions of the heavy atoms using the peak data of dataset VI and these positions were confirmed by the SAD dataset using SOLVE. The initial graphical inspection of the heavy atom sites from SHELX revealed two constellations of 3 Se atoms. The transformation matrix between these constellations was later used in NCS averaging. SHELXE was then used to determine the hand. The final output was a contrast of 0.403 and connectivity of 0.877 for one hand or a contrast of 0.314 and a connectivity of 0.845 for the other, which is a clear discrimination and a contrast of 0.403 is within the

range found for SAD structures (http://biop.ox.ac.uk/www/shelx/shelx-de.pdf). SOLVE was unable to independently identify the positions of all six heavy atoms and so the phases from SHELXD/E were used as the starting point for map generation.

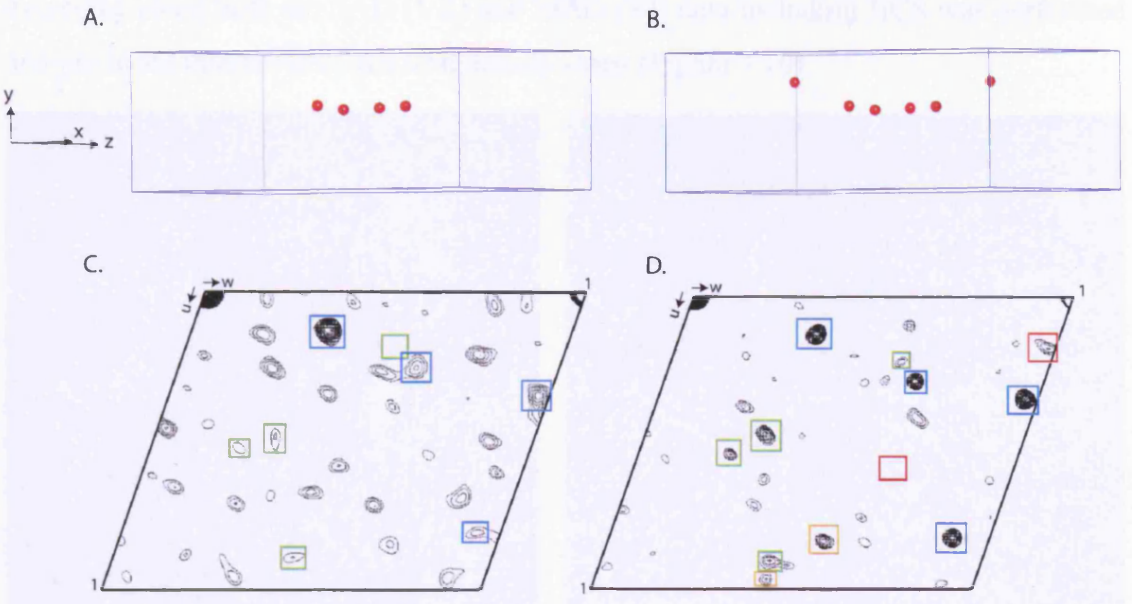

Figure 3.9 – Comparison between wild type and L7M mutant

A and B, unit cell shown in blue and selenium sites represented as red spheres produced using PyMol. A) native protein, B) mutated L7M protein. The original selenium sites are all on the same y-plane; the introduction of additional methionine residues disrupted this arrangement and enabled the phase determination to proceed. C and D, Difference Patterson maps at the v =1/2 Harker section of Se-Met PI31 peak wavelength data contoured at 0.5σ cut at 2σ. C) Wild type and D) L7M mutant. The peaks boxed in blue are cross-peaks and in green are self-peaks from native methionine residues. Those boxed in red are self-peaks and in orange are cross-peaks from the mutated residues. Peaks were identified from calculated Patterson maps after the sites were determined using dataset VI. A and C represent dataset V and B and D dataset VI.

3.6 Model building and refinement of PI31-NTD

The initial phases were not of sufficient accuracy to provide fully interpretable electron density maps. The phases were initially improved using density modification with the additional information from the NCS operator between the 2 Se constellations. Both RESOLVE (Terwilliger 2000) and DM (Cowtan 1994) were used separately and the maps compared in O (Jones *et al.* 1991). The map output from RESOLVE displayed more connectivity and was partially interpretable so some secondary structure elements were built in as a C-alpha chain in O. This secondary structure enabled the refinement of the NCS operator and a further improvement in the phase estimate.

In the analyses of the maps only the peak data from dataset VI had contributed to the phase calculation (Figure 3.10a). The program DMMULTI (Cowtan 1994) allows data from more than one crystal to be combined during density modification. Multi-crystal averaging using both the SAD (VII) and MAD (VI) data including NCS was performed and produced interpretable electron density maps (Figure 3.10).

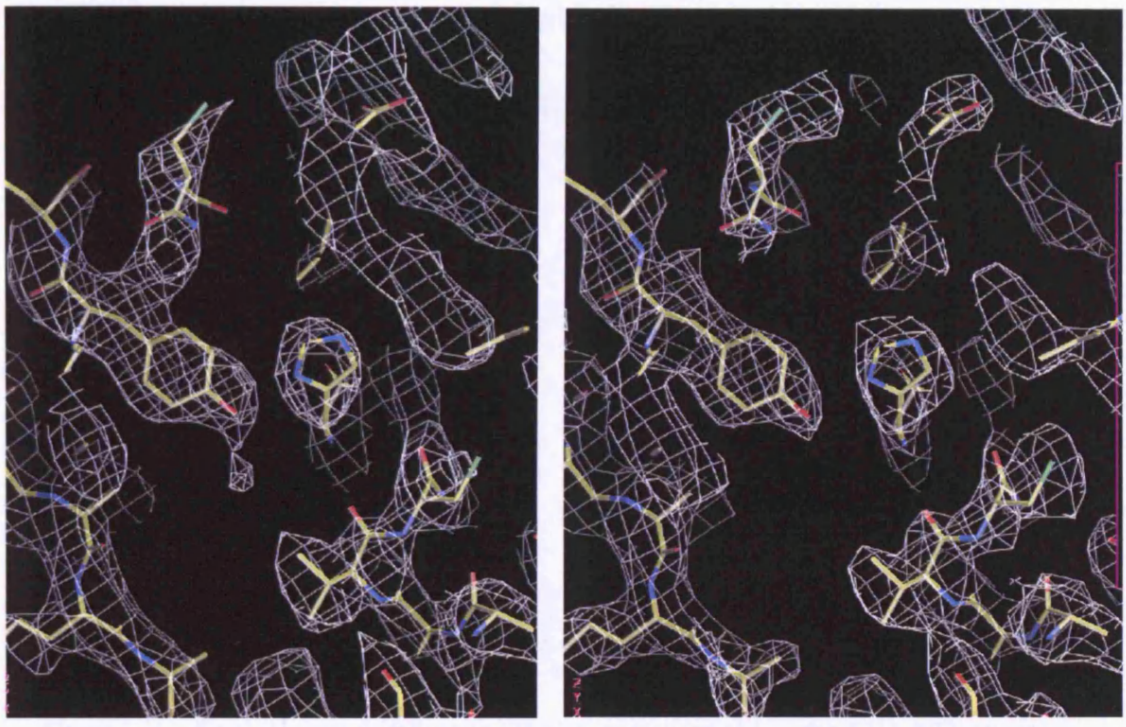

Figure 3.10 – Comparison of electron density maps from RESOLVE and DMMULTI

Electron density maps shown with refined structure. Left) RESOLVE map and Right) DMMULTI map. Unexplained connectivity visible in the RESOLVE map has been solvent flattened in the DMMULTI map enabling correct placement of amino acids.

The positioning of the Se atoms within the anomalous peaks in the electron density assisted chain tracing and the majority of the chain could be assigned in the map from DMMULTI. The program O was used for most of the rebuilding although the final cycles used the newer program COOT (Emsley and Cowtan 2004) as it contained several useful model validation tools.

REFMAC (Murshudov *et al.* 1997) was used for refinement using TLS refinement. Although the L7M mutant does not appear to have an impact on the structure it was thought that the native data would be more valid for future analysis and so the partially refined Se-Met derivatised model was used to solve the native structure by molecular replacement using MOLREP (Vagin and Teplyakov 1997) using the home source native

data (II) (Table 3.3) which is of a comparable resolution to the mutant data. Final data statistics and the structure are those of the native data (dataset II). The structure was refined to 2.64Å and a representative image of the quality of the final electron density map can be seen in Figure 3.11.

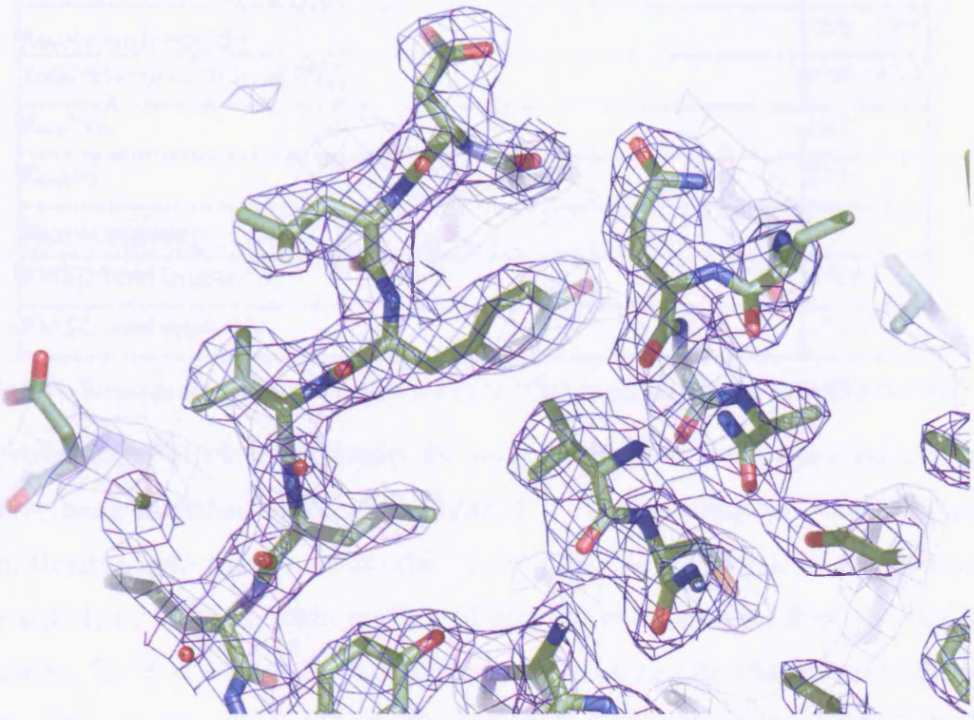

Figure 3.11 – Final electron density map (2Fo-Fc) contoured at 1 σ with PI31-NTD structure

All figures of this type in this thesis were produced using the program PyMOL (DeLano 2004).

3.7 Structure validation

The final structure of PI31-NTD contains the amino acids 1-143 with the remaining 8 amino acids at the C-terminus of the protein disordered and not visible in the electron density. There are also two amino acids on the N-terminus of chain A and one on the N-terminus of chain B that originate from the construct sequence. Whilst they are not part of the PI31-NTD sequence these amino acids are visible in the electron density and so have been built in. The refinement of the structure was concluded when the R_{fact} and R_{free} converged at 19.3% and 27.1% respectively (Table 3.6).

The final model has good stereochemistry (Table 3.6) in that the RMSD values from target bond lengths and from target bond angles are reasonable at 0.016Å and 1.743° respectively.

Model	
Protein atoms present in the structure	2132
Solvent atoms (water oxygen)	39
Protein atoms missing from the structure due to disorder	127
Diffraction Agreement	
Resolution limits (Å)	20.0 - 2.64
Total reflections (test for Rf_{ree})	8707 (476)
R_{fact}(%)	19.3
R_{free}(%)	27.1
Stereochemistry	
RMSD bond lengths (Å)	0.016
RMSD bond angles (°)	1.743

Table 3.6 – Structural quality indicators for PI31-NTD structure output by REFMAC 5.0

During refinement the electron density for several side chains was not complete and so these have been modelled as listed in Table 3.7. The majority of these residues are either in flexible loop regions where the electron density was weak or are intrinsically flexible sidechains at the protein surface. There are two main flexible loop regions in the structure, 72-76 and 95-100. In addition to the missing side-chain density there was no main-chain density for Ser75 in chain A and so this has been omitted; Ser75 is visible in chain B.

Number	Residue	Chain A	Chain B
-1	Ser	CB	Not Built
43	Gln	CB	CB
49	Lys	CB	CB
50	Lys	CB	CB
72	Lys	CB	CA
73	Asp	CB	CA
75	Ser	Not Built	CA
76	Arg	CB	CB
77	Lys	CB	CB
95	Glu		CD
96	Tyr	CB	CB
98	Ser	CB	CB
99	Gln	CB	CB
100	Gln	CB	CB
122	Arg		CB
128	Glu	CB	CB
132	Ser	CB	
142	Ile	CB	CB
143	His		CB

Table 3.7 - Table listing omitted and truncated amino acids

Figure 3.12 shows the Ramachandran plot calculated for the PI31-NTD final structure. This provides a measure of the quality of the proteins main-chain geometry with 88.4% of residues in the most favoured regions, 10.0% in additionally allowed regions, 0.8% in generously allowed regions and 0.8% in disallowed regions. The value within the most favoured regions is above the typical value for this resolution (74.0%) and so this is considered to be a good model. The two residues within the disallowed regions are both in chain A and are Glu86 and Ser98. Whilst Glu86 in both chains has an unusual stereochemistry it is however built into well-defined electron density. Ser98 is in an area of poor electron density and so it is very possible that this geometry suffers from insufficient knowledge to fully place this residue.

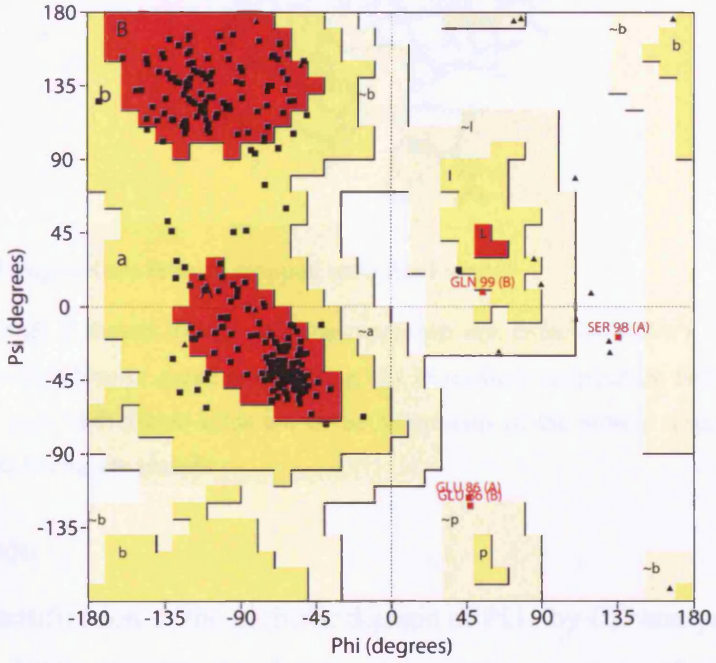

Figure 3.12 – Ramachandran plot of the refined PI31-NTD structure

This plot is calculated using the program PROCHECK (Laskowski *et al.* 1993) using the final refined co-ordinates of the PI31-NTD structure; data from both protomers is shown. Glycine residues are shown as black triangles.

The mean temperature factor for the model was given as 32.2Å^2, which is much lower than the value originally given in the Wilson plot produced by TRUNCATE. Since the temperature factor from Refmac and TLSANL does not include the contribution made by the TLS parameters it is not a directly comparable figure with the Wilson plot. There is a wide range of temperature factors for different residues throughout the structure as can be seen in Figure 3.13 but nevertheless the majority are within a small

range. The residues with high temperature factors (60Å^2) are all from the vector sequence and so are not part of the PI31 protein.

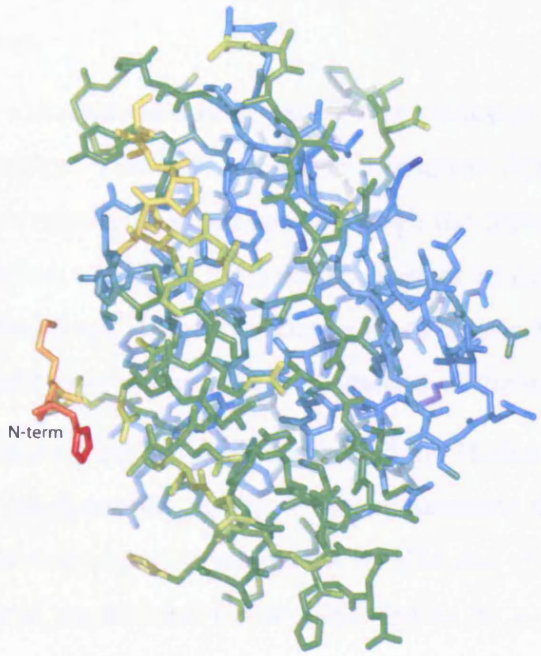

Figure 3.13 – Temperature factors mapped onto final model

Red represents high B-factors (60Å^2) and blue represents low B-factors (20Å^2). The majority of the structure is either blue or green representing low to medium temperature factors as would be expected. The high (red) amino-acids are at the N-terminus of the protein where the additional residues from the vector are present.

3.8 Discussion

The previous identification of the globular domain of PI31 by CD analysis (McCutchen-Maloney *et al.* 2000) enabled the design of a protein construct that produced large quantities of purified recombinant protein suitable for crystallisation. PI31-NTD crystallised readily and after some optimisations diffracted to 2.64Å. This diffraction was highly mosaic, probably due to mechanical weakness of the plate shaped crystals. The high sensitivity to air may also have been due to this morphology, weak inter-molecular contacts within the crystal or the solvent content of the crystal. The cryo protectant used had an effect on the crystal form and so it is likely that this would have had an effect on the internal order of the crystal and so on the resolution and mosaicity. The combination of the high mosaicity and the crystal handling procedures adopted to avoid contact with the air may have reduced the quality of the data collected. In hindsight it would have been preferable to attempt to refine some of the other PI31-

NTD crystallisation conditions to attempt to minimise the mosaicity and so improve the data quality. It is also possible that some of the other conditions may have produced crystals that diffracted to a higher resolution and may have been more amenable to heavy atom derivatisation.

The phasing problems encountered during the structure solution of PI31-NTD required the use of several strategies. The heavy atom derivatisation would, if successful, have supplied the information required for phasing, although the presence of the ammonium iodide in the crystallisation condition reduced the number of derivatives available. In the final structure solution some of the cysteine residues demonstrate a small amount of modification, which could have prevented the interaction of the mercury derivatives.

The selenomethionine protein crystallised readily and the Harker section showed peaks that should have been interpretable and so it was unfortunate that the selenium atom substructure was in a centrosymmetric arrangement. The use of MAD phasing and the introduction of additional methionine residues proved to be a successful method for determining the phases in this situation. It is possible that despite the centrosymmetric arrangement the phases could have been solved using MAD phasing alone, however it is likely that the phase information obtained using this method alone would not have been as good and it still may have been difficult to identify the positions of the selenium substructure.

Despite the problems encountered the final structure is a reasonable model for PI31-NTD as analysed using PROCHECK (Laskowski *et al.* 1993) and when using the R factors for analysis. The analysis with PROCHECK revealed that the model is better than average for the resolution. The difference between the two R factors is larger than would be preferable but it is within the limits for data within the resolution range of 2.5Å to 3Å as set out by Tickle *et al.* (Tickle *et al.* 2000). The other data collected remains to be refined and a higher resolution dataset may reveal further information about the structure of PI31-NTD.

4 Structural and biophysical characterisation of the FP domain

The structure of PI31-NTD is discussed in this chapter along with the sequence and structure based analysis of the previously uncharacterised globular domain in Fbxo7. PI31 and Fbxo7 have a common domain that we have called the FP domain, which spans residues 1-151 in PI31 and 169-276 in Fbxo7. This domain has a novel α/β fold that does not belong to any previously identified family of proteins. The PI31 FP domain forms a dimer through an alpha helix interface as identified in this chapter by gel filtration and AUC of wild type and mutant proteins. Experiments described in this chapter are outlined in a flowchart in Figure 4.1.

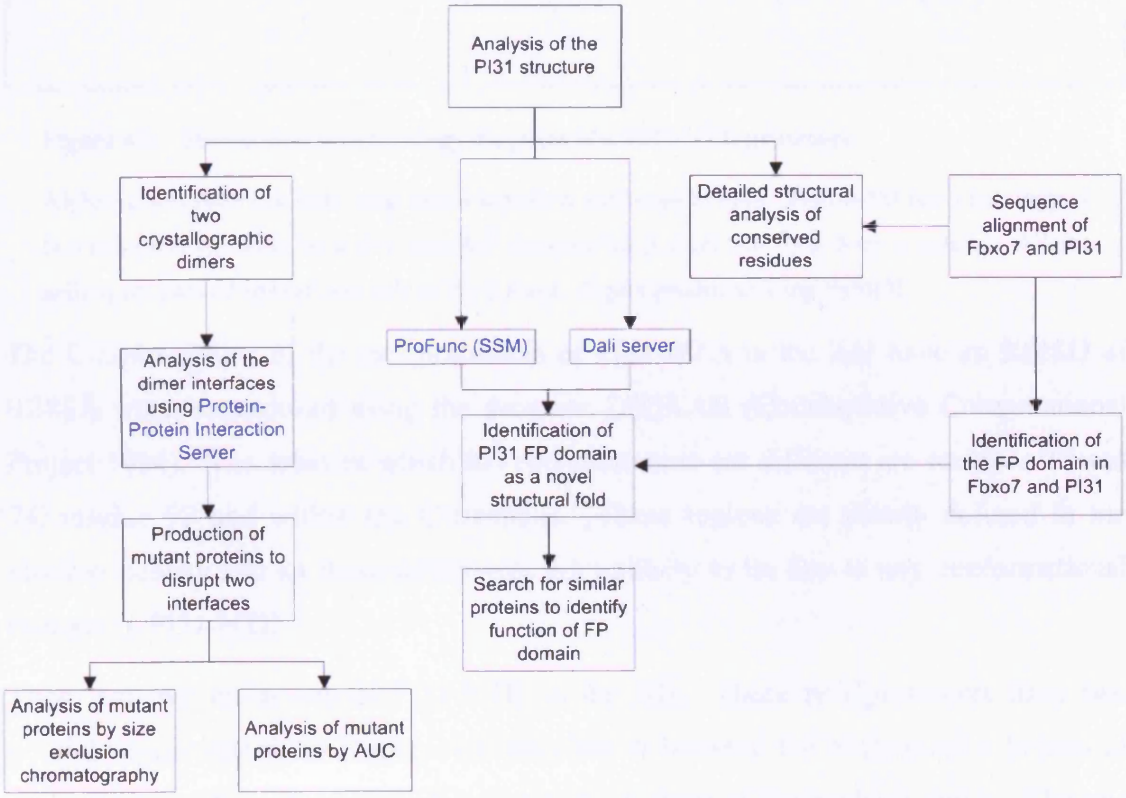

Figure 4.1 – Flowchart of experiments described in chapter four

4.1 Discussion of the structure of the PI31-NTD

The structure of PI31-NTD consists of an α/β fold (Figure 4.2) with the topology of two α helices followed by a five stranded anti-parallel β sheet and then three α helices. All the helices are packed against one side of the β sheet.

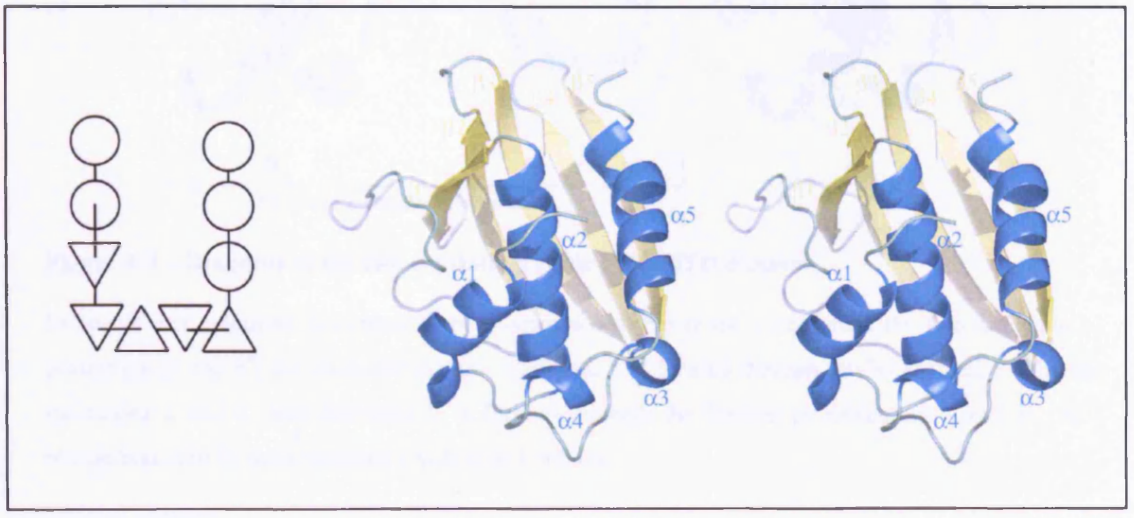

Figure 4.2 – Stereo view and topology diagram of a PI31-NTD protomer

Alpha helices shown in blue, beta sheets in yellow and loops in cyan. PI31-NTD has a topology of two α helices followed by a five stranded anti-parallel β sheet and then three α helices. All the helices are packed against one side of the β sheet. Figure produced using PyMOL.

The C-alpha chains of the two protomers of PI31-NTD in the AU have an RMSD of 0.381Å when superposed using the program LSQKAB (Collaborative Computational Project 1994). The areas in which the two protomers are different are residues 73 and 74, residue 97 and within the C-terminus. These regions are poorly defined in the electron density and so these differences are unlikely to be due to any conformational changes in PI31-NTD.

There are two molecules of PI31-NTD in the AU. These two protomers have two possible dimer interfaces (Figure 4.3) interface A between the N-terminal α helices of molecule a and b' and interface B between the β-sheet of molecules a and b. The two possible dimers are not formed through crystal contacts and so one of these may represent a true dimer in solution.

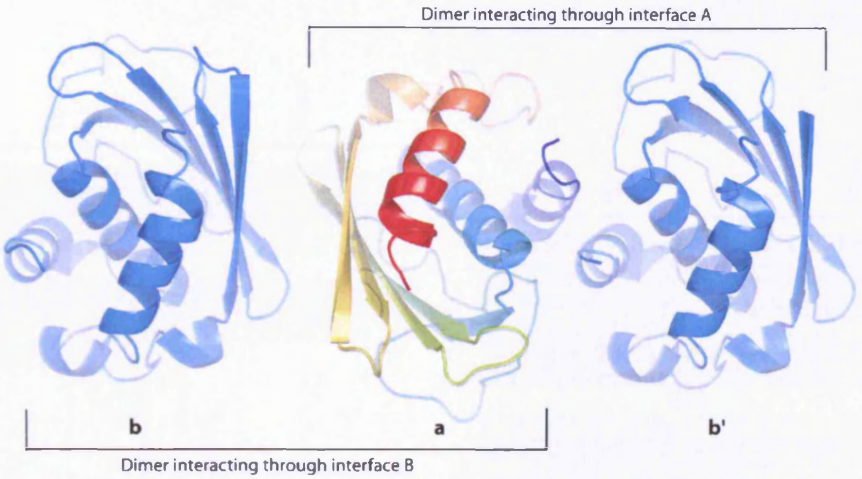

Figure 4.3 – Diagram of the two crystallographic PI31-NTD dimers

Molecule a is coloured from blue at the N-terminus to red at the C-terminus, the two molecule b protomers (b and b') are coloured in blue. Interface A is formed through the N-terminal α helix of molecules a and b' and Interface B is formed through the β-sheet of molecules a and b. No interactions can be seen between a with a or b with b.

4.2 Structure-based alignment of the FP domain

A PSI-BLAST search with PI31-NTD (Altschul *et al.* 1997) identifies Fbxo7 to have sequence similarity. Human PI31-NTD and Fbxo7(169-276) have a 24% sequence identity and 45% sequence similarity. Further analysis using ClustalX (Thompson *et al.* 1997) revealed that PI31 and Fbxo7 contain a common domain that is unique to these two proteins and is conserved throughout all organisms where these proteins are present (Figure 4.4). This previously undefined domain spans residues 1-151 in human PI31 and residues 169-276 in human Fbxo7 and we have termed it the FP domain for **F**bxo7 and **P**I31. This domain was delineated in chapter 2 as being an uncharacterised globular domain in Fbxo7 by limited proteolysis and biomolecular modelling.

From analysis of the sequence alignments it was observed that PI31-NTD contains a conserved YxLxY motif on the β2 strand. In addition to this motif it was also noticed by sequence alignment using ClustalX (Thompson *et al.* 1997) that Asp20 and His28 on the α2 helix were completely conserved through all PI31 proteins indicating a possible enzymatic activity for PI31. All of these residues are buried and form hydrogen bonds as can be seen in Figure 4.5 and so they are unlikely to be functionally important residues within an enzyme active site although they are important for the conformation of the protein.

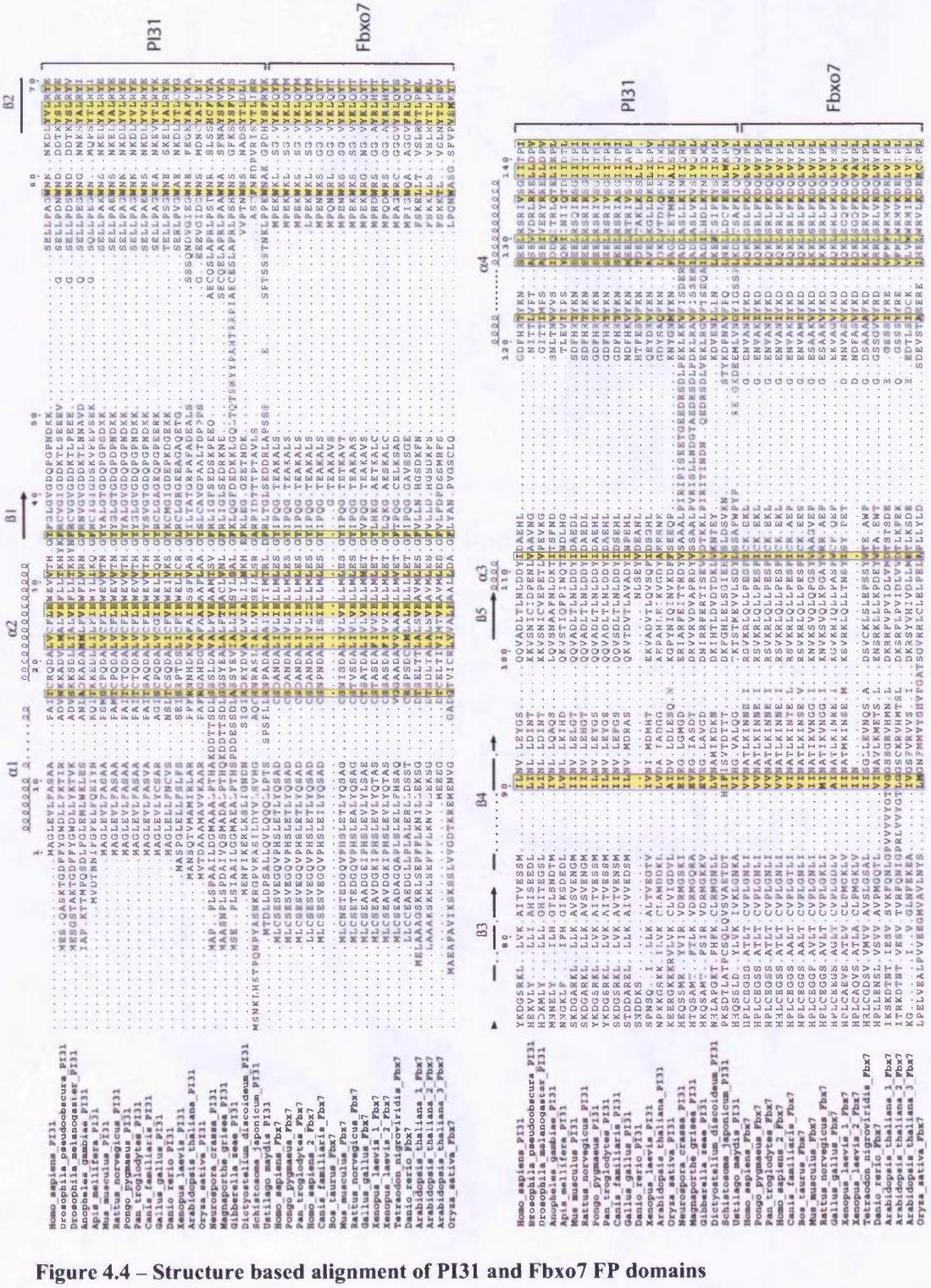

Figure 4.4 – Structure based alignment of PI31 and Fbxo7 FP domains

Structural assignment is based on PI31 FP domain structure. Residues coloured in yellow are conserved, figure produced using the web server ESPript (Gouet *et al.* 1999).

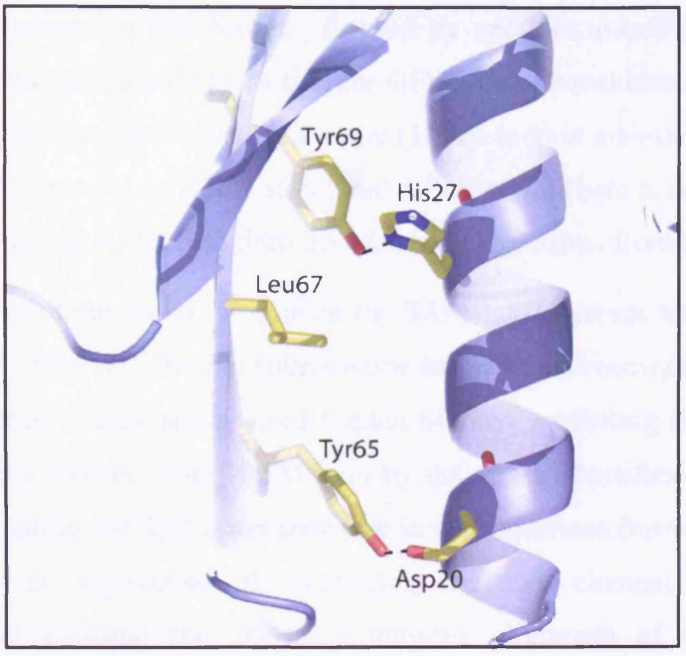

Figure 4.5 – Interaction of the invariant YxLxY motif with Asp20 and His28 FP domain residues

The YxLxY motif on the β2 strand interacts through hydrogen bonds with the conserved residues Asp20 and His28 on the α2 helix. These conserved residues appear to be important in the structural arrangement of the FP domain but are not likely to be functionally active residues.

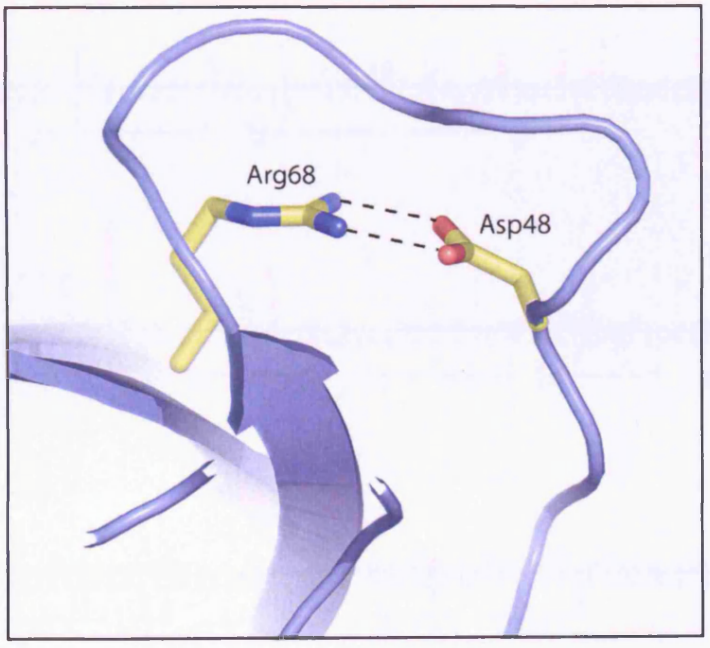

Figure 4.6 – Salt bridge formed by Asp48 and Arg68

Asp48 and Arg68 form a salt bridge in the PI31 FP domain, which stabilises a proline rich loop region.

PI31-NTD also contains a salt bridge formed by residues Asp48 and Arg 68 that stabilises a loop from residue 40 to 64 (Figure 4.6). The two residues are not conserved throughout PI31 proteins and so it is not known if this loop is maintained in a different way for other PI31 proteins or if it is structured differently. There is no known function for this loop and so it is difficult to draw any further conclusions from this interaction.

After identification of the novel FP domain the SAM-HMM server was then applied to further analyse the PI31 FP domain (http://www.cse.ucsc.edu/research/compbio/HMM-apps/T02-query.html). This server used hidden Markov modelling (HMM) to build a profile of the query protein. The HMM built by the server identifies a set of positions that describe the conserved first-order structure in the sequences from a given family of proteins; this is the equivalent of identifying the core elements of homologous molecules. After building this model a multiple alignment of the sequences is performed. The server uses variable, position dependent, gap penalties to improve the quality of the alignment. The model is then used for database searching to identify family members and finally the model is studied directly (Krogh *et al.* 1994).

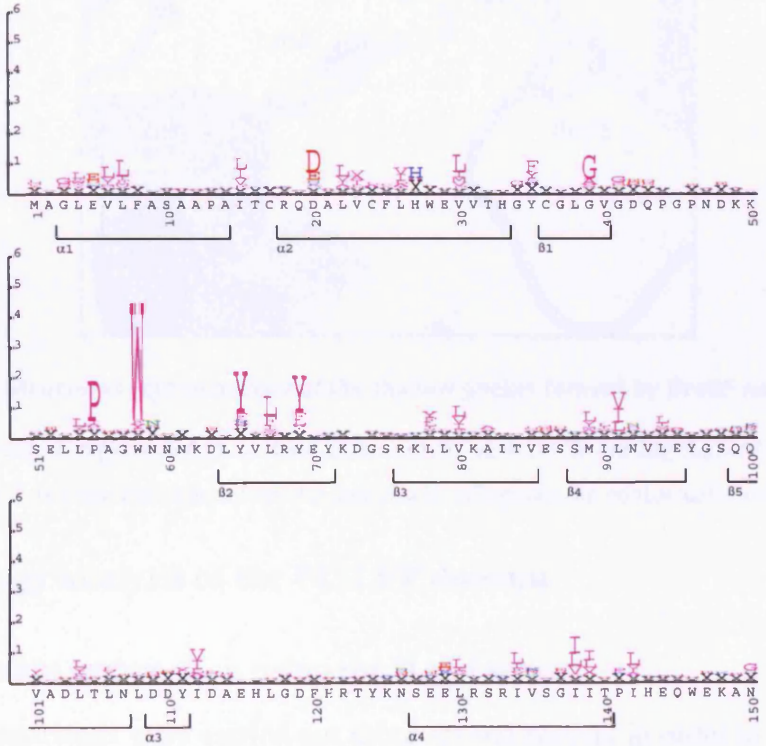

Figure 4.7 – Results from HMM modelling alignment

HMM modelling using PI31-NTD sequence as prey. Size of amino acid indicates the degree of conservation in that position. Secondary structure assignments taken from the PI31 FP domain structure.

PI31 FP domain was applied to the SAM-HMM server in order to identify conserved amino acids that may have been mis-identified by more conventional sequence alignment software (Figure 4.7).

Interestingly the amino acid with the highest Z score is Trp58 followed by Pro55 and these residues are both in the same loop region of the protein. Unlike other residues implicated in the function of the domain the amino acid Trp58 is solvent accessible and seems to be at the bottom of a very shallow pocket that could possibly be involved in protein binding. Pro55 is not solvent exposed but it brings about the change in direction of the peptide backbone that allows the pocket to be formed and so this amino acid could also be important for the function of this protein (Schultz and Schirmer 1990).

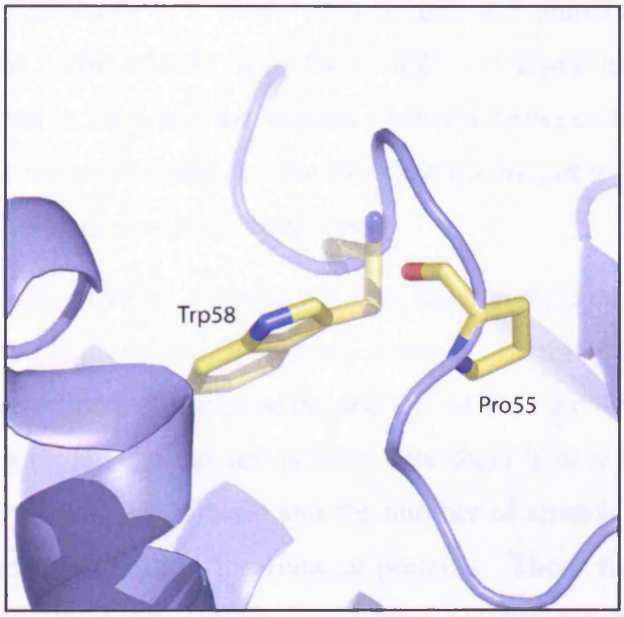

Figure 4.8 – Structural representation of the shallow pocket formed by Pro55 and Trp58

PI31 FP domain is represented in cartoon form. Pro55 and Trp58 are highlighted and it can be seen that Trp58 is a solvent exposed residue and Pro55 influences the conformation of this loop.

4.3 Topology analysis of the PI31 FP domain

4.3.1 Structure comparison using the DALI server

Topology comparisons were carried out using several servers in order to use the 2.64Å structure of the PI31 FP domain to obtain a greater understanding of the function of the PI31 protein. Structural comparison servers can be used to find structural similarity between proteins with little or no sequence similarity and can sometimes help with the elucidation of a function for a previously uncharacterised protein. This method is

becoming more powerful as the algorithms used improve and the number of proteins of known structure in the database increases (Holm and Sander 1994).

The DALI server at the EBI (http://www.ebi.ac.uk/dali/index.html) did not reveal any proteins that have the same topology as the PI31-NTD protomer although one protein was revealed to have a similar topology, the *E. coli* protein CyaY (Cho *et al.* 2000) (pdb code: 1EW4) which is believed to be a closely related ancestor of frataxin (Figure 4.9c). The C-terminal domain of human frataxin was solved by NMR (Musco *et al.* 2000) (pdb code: 1DLX) and revealed to have the same topology as CyaY (Cho *et al.* 2000) (Figure 4.9b). When comparing the topology of PI31-NTD with CyaY using the DALI server it gives a Z score of 4.5. When PI31 FP domain is compared to itself it has a Z score of 32.6 and conversely a Z score of less than 2.0 indicates the proteins are structurally dissimilar. The RMSD fit of the C-alphas of CyaY and PI31 FP domain used in the comparison is 3.4Å and the sequence identity between PI31 FP domain and CyaY is 12%. These values indicate that the two proteins are not topologically identical but that there might be some structural similarity.

All three of the proteins have an anti-parallel beta sheet against one side of which are packed alpha helices (Figure 4.9). There are however several important differences between the structure of the frataxin proteins and that of PI31 FP domain. The number of alpha helices packing against the anti-parallel beta sheet is only two in the frataxins whilst there are five in PI31 FP domain and the number of strands in the beta sheet is five in PI31 FP domain and six in the frataxin proteins. These two large differences mean that the fold of PI31 FP domain is not topologically the same as that of the frataxin proteins despite the similarities. The similarities between the structures of PI31 FP domain and the frataxin proteins could have been used to obtain some information indicating the function of PI31 FP domain by comparison with the function of frataxin. Unfortunately human frataxin, whilst implicated in the disease Friedreich's ataxia, has no reported function (Musco *et al.* 2000).

4.3.2 Structure analysis using the ProFunc server

ProFunc (http://www.ebi.ac.uk/thornton-srv/databases/ProFunc) is a server that aims to predict the function of a protein of known structure by using both sequence and structural information to identify functional motifs or close relationships to previously characterised proteins (Laskowski *et al.* 2005).

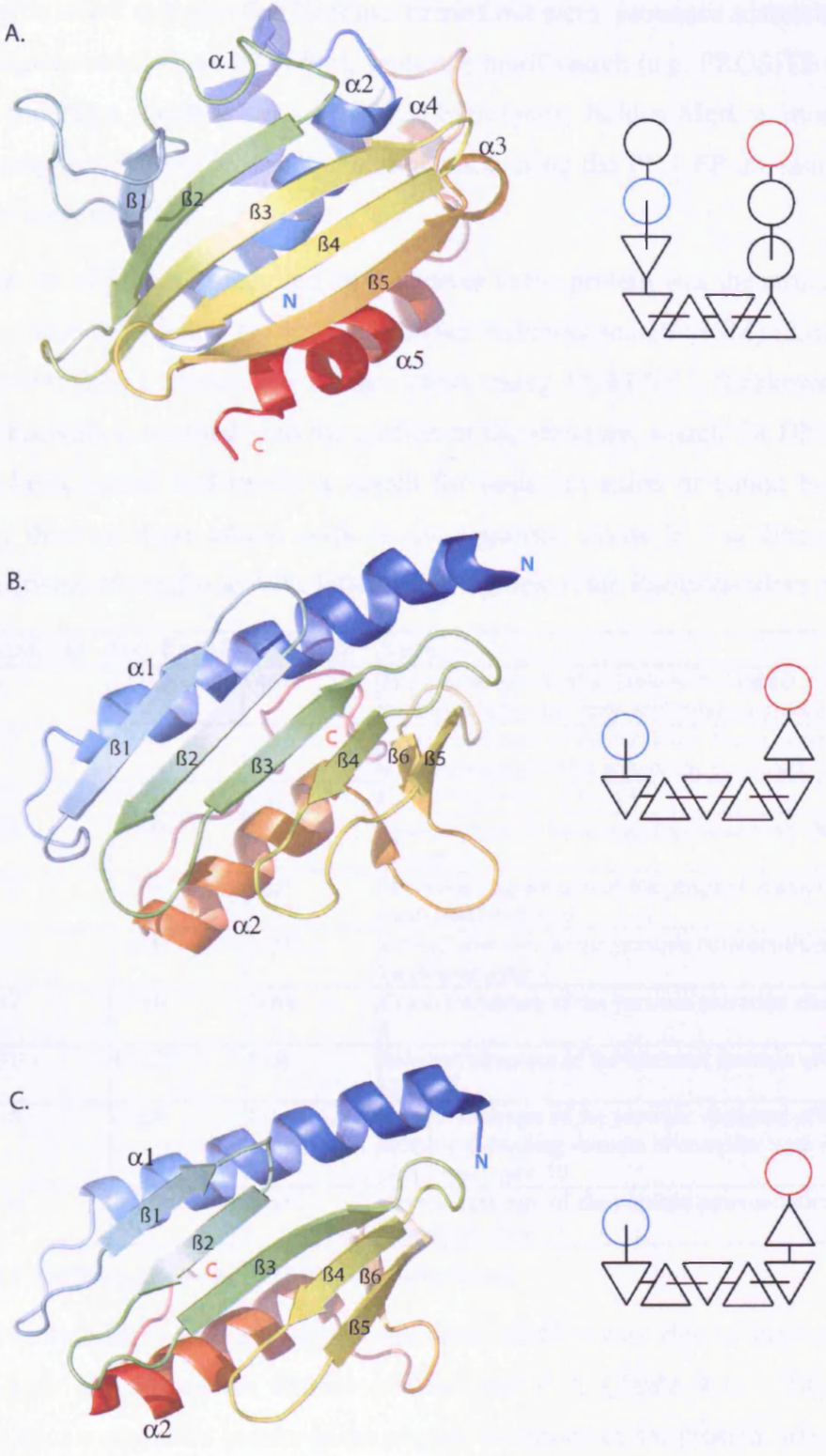

Figure 4.9 – Cartoon representation of PI31 FP domain, frataxin and CyaY for topological comparison

A. PI31 FP domain protomer crystal structure. B. Frataxin C-terminal domain NMR structure. C. CyaY crystal structure. All structures have a topology diagram to the right of the structure and are coloured from blue at the amino terminus to red at the carboxy terminus.

The sequence based analyses that ProFunc carried out were: sequence search against the PDB, sequence search against UniProt, sequence motif search (e.g. PROSITE (Sigrist *et al.* 2002) and Pfam (Bateman *et al.* 2004)), superfamily hidden Markov model search and gene neighbours search. In all analyses made using the PI31 FP domain sequence no proteins were reported.

The second set of processes applied by the server to the protein was the structure based analyses. These included: a secondary structure matching search (SSM) (Krissinel and Henrick 2004, 2005), search for surface clefts using SURFNET (Laskowski 1995), residue conservation mapped onto the surface of the structure, search for DNA-binding helix-turn-helix motifs and finally a search for nests (an anion or cation binding site formed by three or more amino acids in the sequence whose Ψ – φ dihedral angles alternate between the right- and the left-handed regions of the Ramachandran plot).

Z-Score	RMSD (Å)	Seq. ID (%)	PDB entry	Name
4.4	2.1	17.90	1sa8	The nmr structure of a stable and compact all-beta-sheet variant of intestinal fatty acid-binding protein 2
3.3	2.79	18.90	1xkp	Crystal structure of the virulence factor yopn in complex with its heterodimeric chaperone sycn-yscb 3
2.5	3.02	17.00	1jya	Crystal structure of syce 4
2.5	3.22	17.90	1ry9	Spa15, a type iii secretion chaperone from shigella flexneri 5
2.5	2.73	12.50	1kaf	DNA binding domain of the phage t4 transcription factor mota (aa105-211) 6
2.5	2.92	18.90	1n5b	Crystal structure of the yersinia enterocolitica molecular chaperone syce 7
2.4	2.97	17.90	1k6z	Crystal structure of the yersinia secretion chaperone syce 8
2.3	3.31	14.20	1soy	Solution structure of the bacterial frataxin orthologue, CyaY 9
2.3	2.98	17.00	1l2w	Crystal structure of the yersinia virulence effector yope chaperone-binding domain in complex with its secretion chaperone, syce 10
2.1	3.16	16.00	1md1	Crystal structure of the yersinia enterocolitica molecular chaperone syce

Table 4.1 – SSM results using PI31 FP domain to search.

The SSM search did reveal several proteins upon which it was able to superimpose the C-alpha chain and obtain an RMSD of less than 4 Å (Table 4.1). This C-alpha alignment allows extensive breaks in the peptide backbone of the protein, which reveals many false positives. For example the protein with the highest Z score (PDB code: 1SA6) was an intestinal fatty acid-binding protein, which consists entirely of beta strands. None of the proteins described in Table 4.1 have the same topology as PI31 FP domain and, as before, the protein with topology most similar is CyaY.

In PI31 FP domain, using the ProFunc server, no surface clefts were identified, there were no patches of conserved residues on the surface, no DNA helix-turn-helix binding motifs and all the nests are predicted to be too shallow to have any functional significance.

The final analyses the server performed on the protein structure was a 3D template search which included: a search for enzyme active sites, ligand binding sites and DNA binding sites. None of these searches produced any significant matches.

The ProFunc server provided a very thorough examination of the PI31 FP domain structure with respect to other previously characterised proteins and it did not produce any significant hits. It appears from the searches using both DALI and ProFunc that PI31 FP domain contains a novel fold and further analysis will be required to elucidate its function.

4.4 Analysis of the two crystallographic dimer forms

As described previously in section 4.1 there were two possible dimer formations within the crystal structure (Figure 4.10). In order to identify the correct physiological dimer a protein-protein interaction server was used to look at the features of the two possible dimer interfaces (http://www.biochem.ucl.ac.uk/bsm/PP/server/) (Jones and Thornton 1995, 1996). The results of the server analysis of both dimer interfaces are shown in Table 4.2.

Protein Interface Parameter	Value interface A	Value interface B
Interface Accessible Surface Area (Å^2)	557.22	279.89
% Interface Accessible Surface Area	7.84	4.04
Planarity	1.56	1.17
Length & Breadth	25.81 & 15.42	25.92 & 10.22
Length/Breadth Ratio	0.69	0.36
Interface Residue Segments	4	3
% Polar Atoms in Interface	29.16	36.01
% Non-Polar Atoms in Interface	70.8	63.9
Secondary Structure	Alpha	Beta
Hydrogen Bonds	5	2
Salt Bridges	0	0
Disulphide Bonds	0	0
Gap Volume	3563.36	3691.66
Gap Volume Index	3.23	6.58

Table 4.2 – Output data from Protein-Protein Interaction Server for the two possible interfaces (Jones and Thornton 1995, 1996)

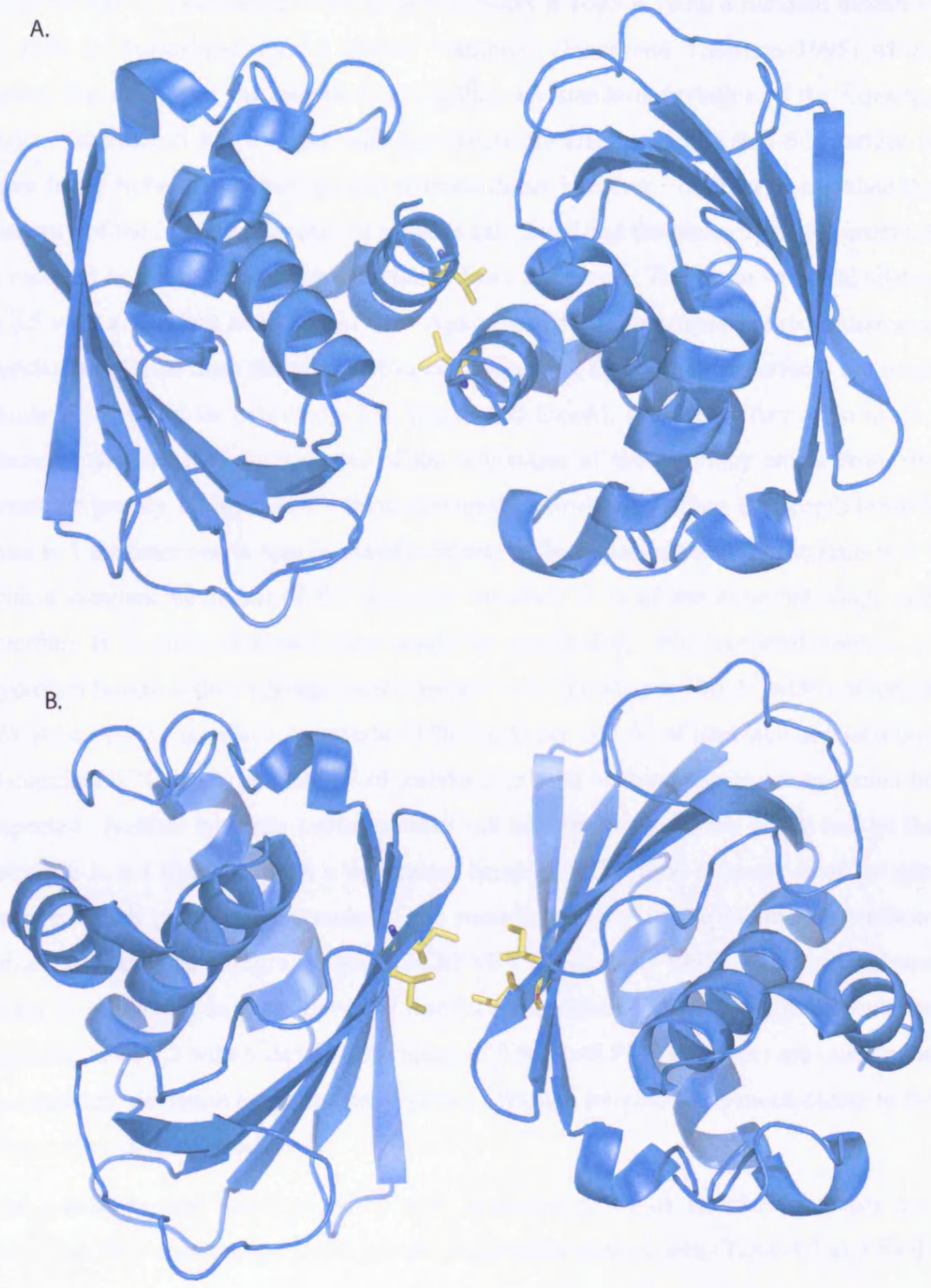

Figure 4.10 – Structural diagram of the two possible dimer interfaces and mutations

A. Interface A shown with the interface residue Val6 highlighted. B. Interface B shown with the residues Ile83 and Ile90 highlighted

The interface accessible area is reported for only one of the protomers and the expected value for typical functionally relevant homodimers is 1685 $Å^2$ with a standard deviation of 1101 $Å^2$ (calculated for 32 known examples) (Jones and Thornton 1995) which means that neither of the interfaces are within one standard deviation of the expected value. Interface A has a larger interface accessible area indicating that this surface is more likely to be the physiologically relevant dimer interface. In order to calculate the planarity of the interface the best fit plane is calculated and the planarity of the interface is reported as a root mean square deviation from this plane. The mean value calculated is 3.5 with a standard deviation of 1.7. Again both of the interfaces are more than one standard deviation from the expected values, indicating that both the interfaces are more planar than would be expected. The length and breadth provides information on the shape of the interface by analysis of the deviations of the interface atoms from the centre of gravity of the best fit plane previously calculated. When the length/breadth ratio is 1 the interface is approximately circular. The expected value of the ratio is 0.7 with a standard deviation of 0.2 therefore interface A is of the expected shape and interface B is more extended than would be anticipated. The expected number of hydrogen bonds within a homodimer interface is 0.7 bonds per 100 $Å^2$ with a standard deviation of 0.5. Interface A contains 0.90 bonds per 100 $Å^2$ of interface and interface B contains 0.71 bonds per 100 $Å^2$ of interface so both of these values are as would be expected. Neither interface contains either salt bridges or disulphide bonds and so the interface is not likely to form a permanent complex. The final measure is of the gap volume which provides a measure of the complementarity of the interacting surfaces which is calculated using a program SURFNET (Laskowski 1995). The gap volume index is calculated as (gap volume / interface accessible surface area) and would be expected to be 2.2 with a standard deviation of 0.9. Both PI31 interfaces are outside the one standard deviation range for this measure although interface A is much closer to the expected value than interface B.

The protein-protein interface server also produced a list of interface residues and calculates their contribution to the interface accessible surface area (Table 4.3 and Table 4.4).

In the case of interface A the major contributor to the accessible surface area is Val6 and so this was chosen as the optimal residue to target to disrupt this dimer interface (Figure 4.10a). For interface B Ile83, Val85, Glu86 and Ile90 contribute significantly to

the accessible surface area of the interface. Glu86 is at the edge of the accessible surface area and so mutation of this amino acid would generate data that was difficult to analyse when looking to disrupt the dimer interface. Analysis of the structure revealed that Ile83 and Ile90 were both at the centre of the dimer interface and were positioned in different areas of the dimer interface and so these two amino acids were chosen for mutation to disrupt interface B (Figure 4.10b).

Residue Number	Residue Name	Interface ASA	% Interface ASA
2	ALA	54.14	9.72
3	GLY	27.67	4.97
5	GLU	72.65	13.05
6	***VAL***	***109.07***	***19.59***
7	LEU	5.1	0.92
9	ALA	34.22	6.15
10	SER	31.53	5.66
28	TYP	27.32	4.91
31	VAL	3.78	0.68
32	THR	12.92	2.32
50	LYS	14.39	2.58
51	SER	13.92	2.5
52	GLU	71.96	12.92
53	LEU	60.74	10.91
121	HIS	17.45	3.13

Table 4.3 – Analysis of interface by residue for interface A

Residues to be mutated are highlighted in bold and italics

Residue Number	Residue Name	Interface ASA	% Interface ASA
42	ASP	6.64	2.38
64	LEU	39.2	14.03
66	VAL	5.83	2.09
81	LYS	24.98	8.94
83	***ILE***	***49.17***	***17.6***
85	VAL	48.17	17.24
86	GLU	64.82	23.2
90	***ILE***	***39.07***	***13.99***
92	ASN	1.5	0.54

Table 4.4 - Analysis of interface by residue for interface B

Residues to be mutated are highlighted in bold and italics.

The residues identified as contributing to the dimer interface are all hydrophobic residues. In accordance with this the analysis of the electronic surface potential of each of the dimer interfaces shows that the surfaces of both of the possible dimer interfaces of the PI31 FP domain are hydrophobic (Figure 4.11). There are no areas of charged amino acids on the surface of the interfaces and so the most efficient way to disrupt the interface was to introduce a charged residue in to the interface and use the repulsion of the charge to disrupt the dimerisation of the PI31 FP domain. The mutation made in the FP domain to disrupt interface A consisted of mutating residue 6 from valine to arginine, which will be referred to as V6R. A double mutation was made to disrupt interface B consisting of mutating both residues 83 and 90 from isoleucine to glutamic acid, which will be referred to as I83,90E. The mutants were designed to have the maximum effect on the dimer interface but with minimum number of basepairs altered in the triplet code in the DNA sequence. All mutations were prepared using the Quikchange protocol and primers described in appendix A.

4.5 Biophysical characterisation of the PI31 FP domain

4.5.1 Analytical size exclusion chromatography of PI31 FP domain

Expression of the mutant PI31 FP domain proteins was carried out as described in appendix A and purified as described for His-tagged proteins. The mutant proteins expressed at similar levels to the wild type protein (Figure 4.12) but the introduced mutations I83E and I83,90E cause atypical behaviour of the proteins when analysed by SDS-PAGE. This difference can also been seen when the protein is expressed in mammalian cells (Heike Laman). The change in charge on the surface of the protein may cause the protein to appear to be larger than expected as SDS-PAGE is affected not only by the size and shape of the protein but also by its overall charge. The calculated molecular weight for all the PI31 FP domain proteins is 16.9kDa and both the wild type protein and V6R mutant have an apparent size of around 17kDa on SDS-PAGE. The I83E mutation gives the protein an apparent size of 18kDa and the I83,90E double mutation an apparent size of just under 20kDa. The pI of wild type PI31 FP domain is 5.14 and this changes to 5.30 and 4.93 for the V6R and I83,90E mutations respectively (Bjellqvist *et al.* 1993). The change in pI is not large and would be unlikely to have the large effect on the SDS-PAGE seen and so it is possible that this is due to a conformational change affecting the binding of SDS to the protein.

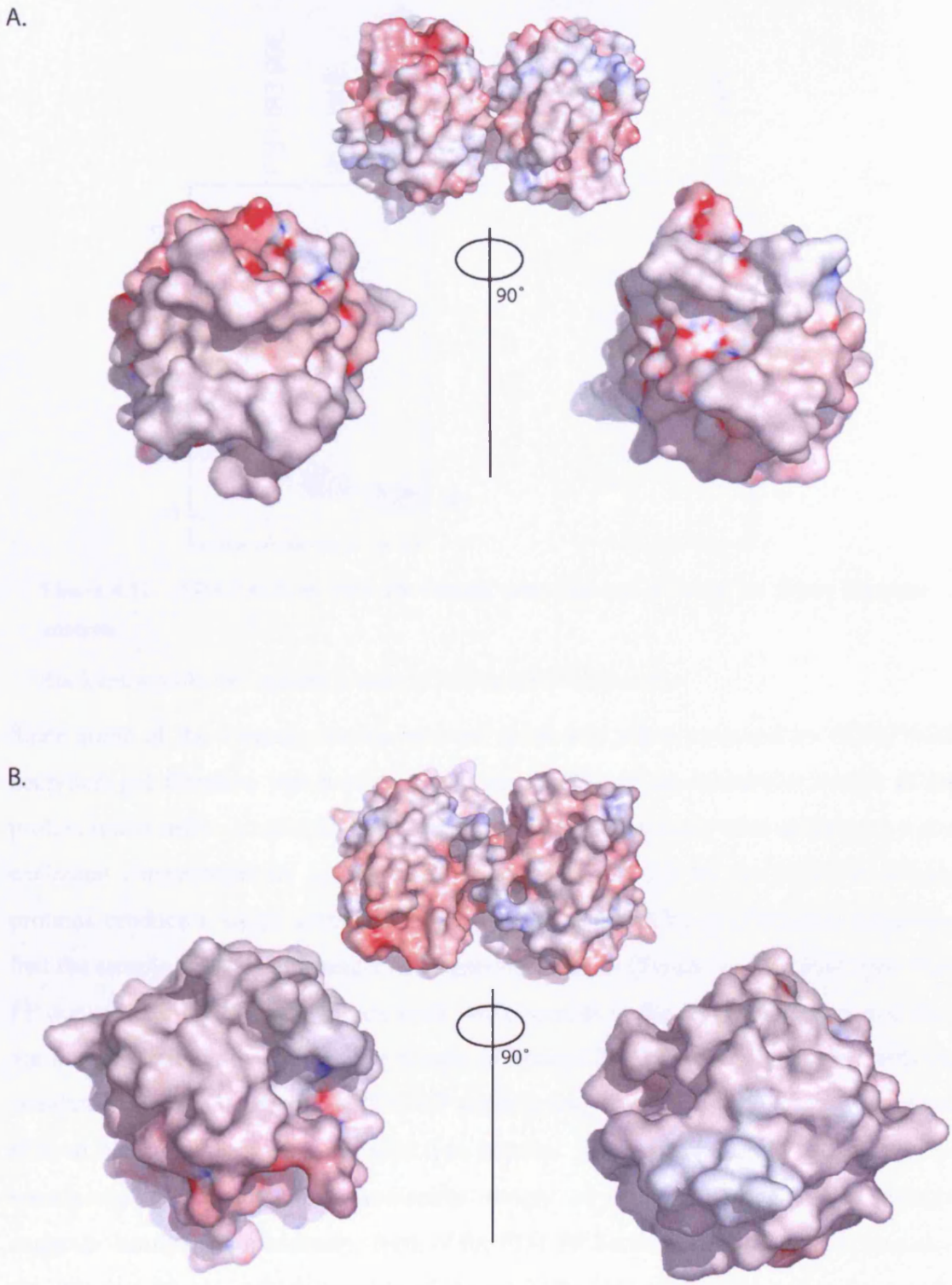

Figure 4.11 – Electrostatic potential of the surface of the PI31 FP domain

A. Electrostatic potential of the surface of PI31 FP domain when dimerised through interface A and the surface of the interface itself. B. Electrostatic potential of the surface of PI31 FP domain when dimerised through interface B and the surface of the interface itself.

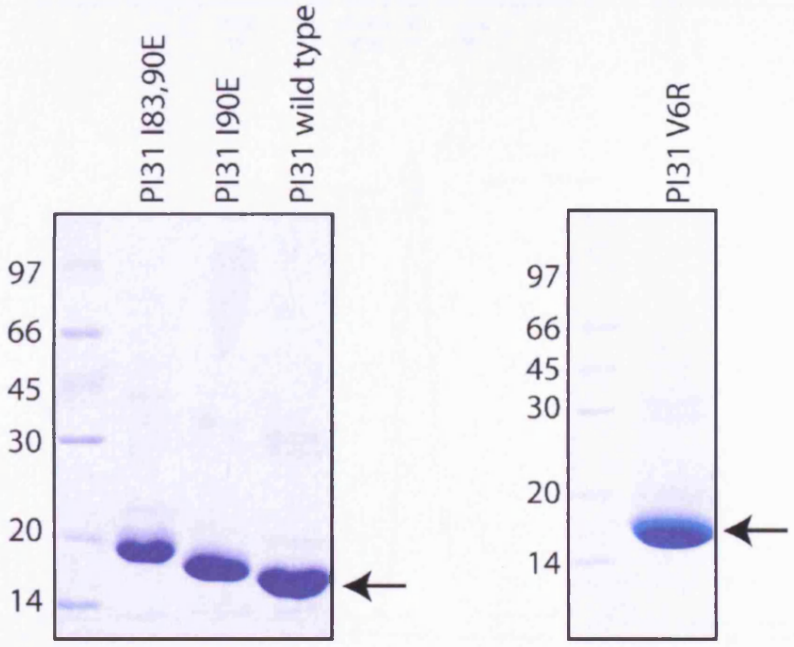

Figure 4.12 – SDS-PAGE of PI31 FP domain wild type and mutants for dimer interface analysis

Black arrow marks the expected position of wild type PI31 FP domain.

Since some of the mutant proteins behaved aberrantly when analysed by SDS-PAGE analytical gel filtration was used to obtain an estimate of the molecular weight of the protein under native conditions. All three of the mutant proteins were analysed by size exclusion chromatography as described in appendix A. All of the PI31 FP domain proteins produce a single symmetrical peak when analysed by gel filtration indicating that the sample consists of a single homogeneous species (Figure 4.13). Wild type PI31 FP domain and I83,90E mutant are both single species under these conditions and they are both dimeric, with a molecular weight of around 31 kDa when compared with the standard proteins. Interestingly PI31 FP domain with the single V6R mutation did not elute at the same position as the wild type protein. The mutant protein appeared to be homogeneous and eluted at a molecular weight of about 15 kDa, which strongly suggests that this is a monomeric form of the PI31 FP domain. Since the V6R mutation disrupts interface A, which consists of the α1 helix, this change in molecular weight indicates that the dimer interface of PI31 FP domain is through the α1 helix of both protomers. Gel filtration analysis is performed on proteins at high concentration so further work was required to fully characterise the possible protein-protein interactions under more physiological conditions.

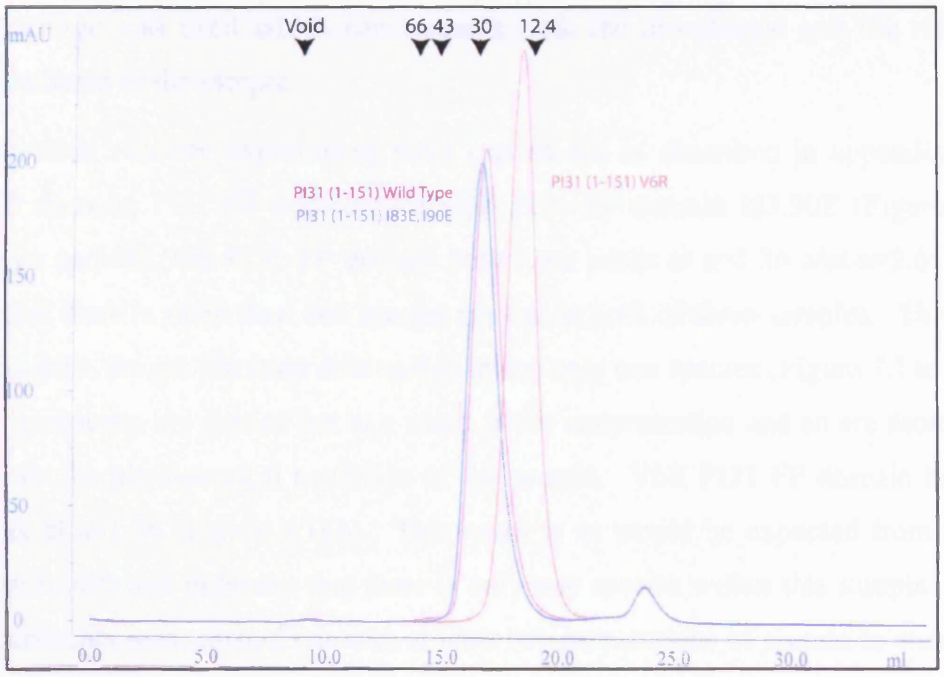

Figure 4.13 – Comparison between wild type and mutant PI31 FP domain by gel filtration

Gel filtration performed using a Superdex 75 HR 10/30 column (Amersham Biosciences) at a rate of 0.4 ml/min using 20mM Tris-HCl pH 8.0, 50mM NaCl as buffer. 500µl of protein sample was loaded onto the column. Superposition of gel filtration traces from PI31 FP domain wild type (pink), I83,90E (blue) and V6R (red). Black arrows show the positions of molecular weight markers (in kDa). y axis is absorbance at 280nm measured in mAU and x axis is volume in ml.

4.5.2 Analytical ultracentrifugation of the PI31 FP domain

In order to establish the behaviour of PI31 FP domain under physiological conditions analytical ultracentrifugation (AUC) was used to analyse both wild type and mutant proteins. AUC is a classical method for the analysis of purified proteins in dilute solution and can be used to characterise the aggregation state, heterogeneity and thermodynamic interactions of proteins. AUC consists of the application of a centrifugal force with simultaneous real-time observations of the distribution of the protein. The resulting data can then be analysed from first principles. The main strength of this method is that it does not require any chemical modification of the protein and since there is no interaction with a matrix the proteins can be studied in solution. AUC experiments utilise the fact that faster sedimenting complexes are transported through a solution of the slower sedimenting components. Reversibly formed complexes that subsequently dissociate can readily re-associate during the experiment, thus permitting the hydrodynamic and thermodynamic characterization of even weak and transient interactions (Scott and Schuck 2005). For these experiments a Beckman XL-I

ultracentrifuge was used which can measure both the absorbance and the refractive index gradients of the sample.

Sedimentation velocity experiments were carried out as described in appendix A for PI31 FP domain, PI31 FP domain V6R and PI31 FP domain I83,90E (Figure 4.14). Wild type and I83,90E PI31 FP domain both have peaks at s=1.36 and s=2.65 which shows that there is more than one species present in both of these samples. This is not expected from the gel filtration data as this shows only one species (Figure 4.14a and c). AUC experiments are carried out at a much lower concentration and so are more likely to indicate the physiological condition of the protein. V6R PI31 FP domain has only one peak at s=1.36 (Figure 4.14b). This result is as would be expected from the gel filtration results and indicates that there is only one species within this sample. All of the experiments were carried out with at least two preparations of protein to ensure that the effects observed were not due to an individual protein preparation. The V6R c(s) distribution has only one peak and so these results were modelled on a c(M) distribution to obtain an estimate of the molecular weight. This is not accurate as the result is dependent on the shape of the protein (the frictional ratio) and this is not known *a priori* for PI31 V6R FP domain. The c(M) distribution gives the molecular weight of PI31 FP domain V6R as 18.1kDa which is higher than predicted from the amino acid sequence (16.9kDa) but it is consistent with this mutation causing the protein to be monomeric. The c(M) distribution was plotted for PI31 FP domain wild type protein, as a comparison, and the value calculated for the main peak gives a molecular weight of 33.1kDa and for the minor peak 14.8kDa which are likely to represent the dimeric and monomeric protein respectively (Figure 4.14d). This result indicates that the wild type protein is in a monomer-dimer equilibrium with the predominant species being dimeric.

Having established that there is more than one species present in both the PI31 FP domain wild type and the I83,90E samples a sedimentation equilibrium run was needed to calculate the thermodynamic properties of these two species at equilibrium. Sedimentation equilibrium experiments were performed as described in appendix A on wild type, V6R and I83,90E PI31 FP domains at each of three concentrations and at three different rotor speeds.

Sedimentation equilibrium data was fitted with a model for self-association using the program SEDPHAT (Laue *et al.* 1992) and from this binding constants were determined. For monomer-dimer self-association, mass action means that the amount

of dimer can be expressed as $c_2=K_{12}c_1^2$ (with concentrations and binding constants expressed in molar units), and the sedimentation equilibrium profile then follows the relationship expressed in Equation 4.1 assuming an ideally sedimenting system.

Equation 4.1 $$a(r) = c_1(r_0)\varepsilon_1 d\exp\left[M_b\frac{\omega^2}{2RT}\left(r^2-r_0^2\right)\right]+K_{12}c_1(r_0)^2\,2\varepsilon_1 d\exp\left[2M_b\frac{\omega^2}{2RT}\left(r^2-r_0^2\right)\right]$$

Where M_b denotes the monomer buoyant molar mass $M=(1-\bar{v}\rho)$ which can be predicted from amino acid composition.

Equation 4.1 was used by the program SEDPHAT to produce predicted sedimentation equilibrium curves based on the known parameters (Figure 4.15). The RMSD of the fit is given in Table 4.5 – RMSD of the sedimentation equilibrium experiment and Kd for self-association

and these show that the fit is very good with a low RMSD for all the samples tested. The residuals in Figure 4.15 show that any errors in the data are small and random and so should not affect the binding constants calculated.

Sedimentation equilibrium analysis revealed that the monomer-dimer equilibrium predicted by the sedimentation velocity experiment is a good fit for both wild type and I83,90E FP domains with a K_d of 4.85μM and 3.95μM respectively. The previously reported monomer-dimer equilibrium K_d for full length protein was 6.25μM (McCutchen-Maloney *et al.* 2000) which is of the same order of magnitude as the values for the FP domain indicating that the dimer interface is predominately through the FP domain. This represents only moderate binding, which is presumably why the protein is in dynamic equilibrium instead of a permanent dimer. Based on the binding constant of 6.25μM McCuthchen-Maloney *et al* (McCutchen-Maloney *et al.* 2000) predicted that in intact cells nearly all PI31 would be monomeric. The monomer-dimer equilibrium model does produce a good fit of the data to PI31 FP domain V6R with the K_d calculated for this protein being 462M. This value is physiologically impossible implying that this protein does not form a dimer in solution, which fits with all other data for this protein presented here.

	Local rmsd 0.5 mg/ml	Local rmsd 0.25 mg/ml	Local rmsd 0.1 mg/ml	K_d
Wild Type	0.010734 (n=184)	0.008667 (n=154)	0.004149 (n=166)	**4.87 μM**
I83,90E	0.022778 (n=229)	0.0345 (n=158)	0.013730 (n=313)	**3.95 μM**
V6R	0.014677 (n=202)	0.008626 (n=190)	0.009948 (n=192)	**462.4 M**

Table 4.5 – RMSD of the sedimentation equilibrium experiment and K_d for self-association

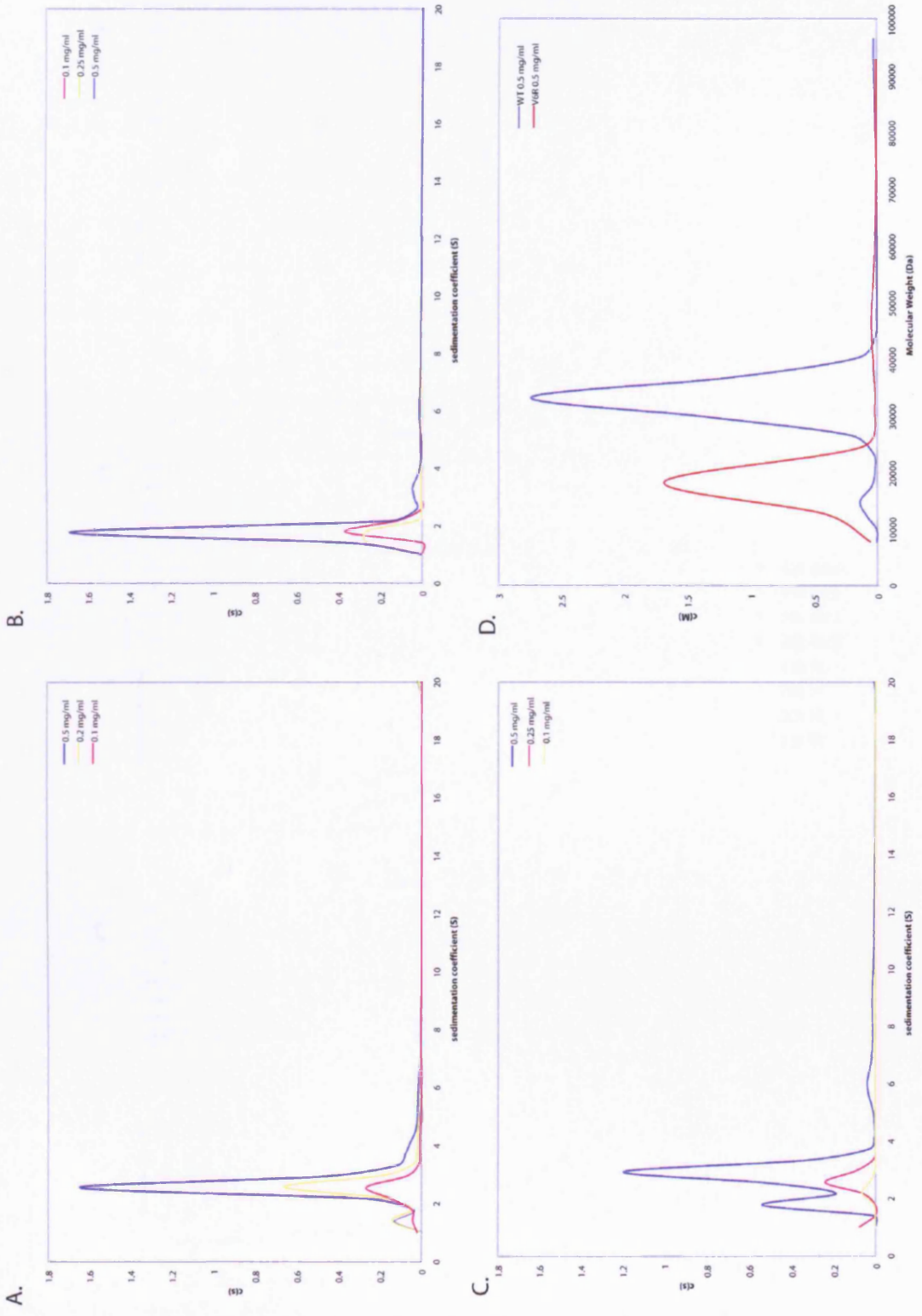

Figure 4.14 – c(s) and c(M) distributions for PI31 FP domain proteins

A. c(s) distribution for PI31 FP domain. B. c(s) distribution for PI31 FP domain V6R. C. c(s) distribution of PI31 FP domain I83,90E (all c(s) distributions shown for three concentrations. D. c(M) distribution of PI31 FP domain (blue) and PI31 FP domain V6R (red) shown for the highest concentration.

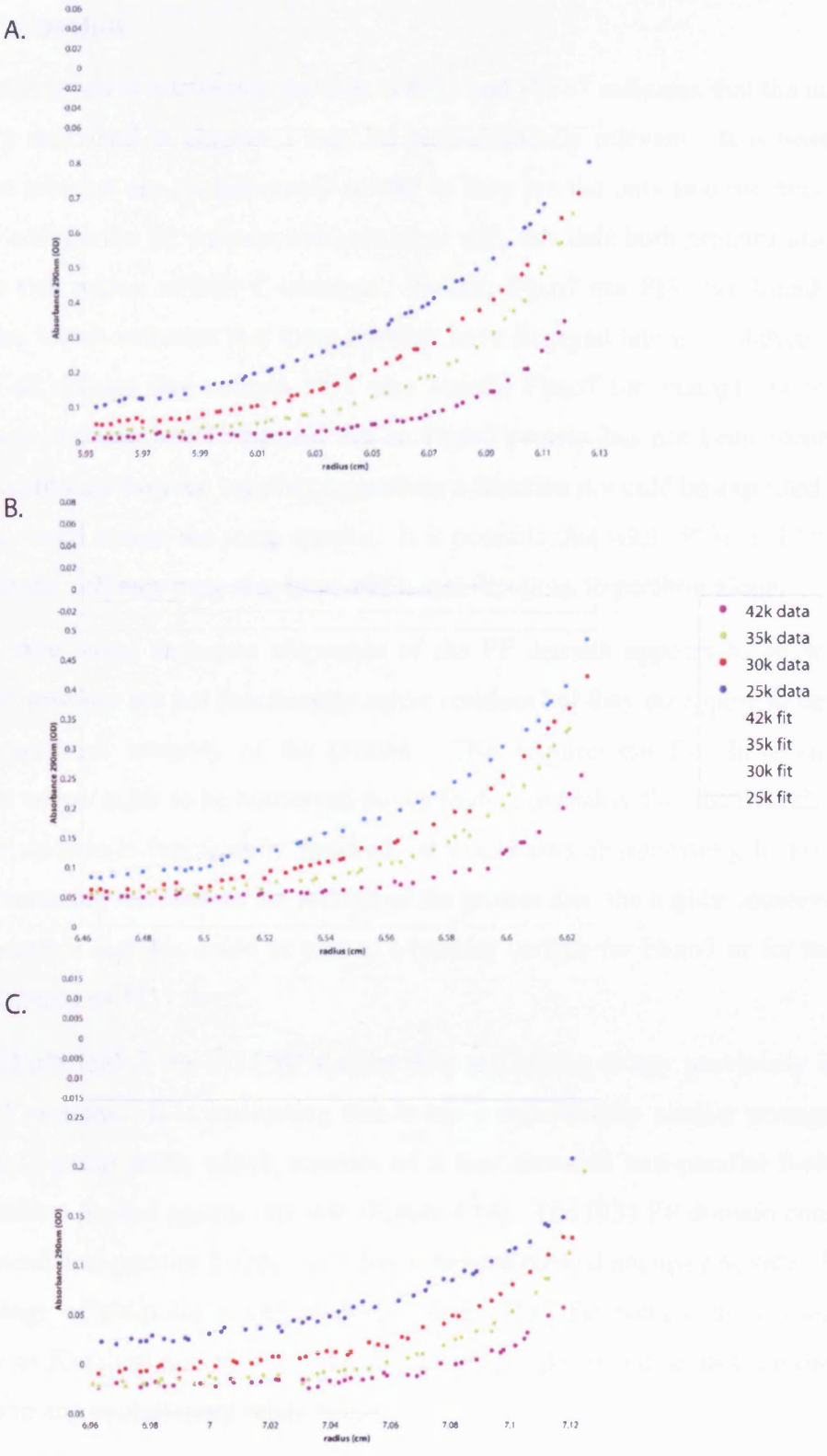

Figure 4.15 – Typical sedimentation equilibrium data for PI31 FP domain proteins

The top box is the residuals and the lower is the experimental profile with the model fit also shown. A. Typical profile for PI31 wild type. B. Typical profile for PI31 I83,90E. C. Typical profile for PI31 V6R.

4.6 Discussion

The identification of a common domain in PI31 and Fbxo7 indicates that the interaction originally described in chapter 2 may be physiologically relevant. It is possible that these two proteins are evolutionarily related as they are the only two proteins we have found to contain the FP domain and consistent with this data both proteins also contain a proline rich region at their C-terminus. Neither Fbxo7 nor PI31 are found in lower eukaryotes which indicates that these proteins have emerged late in evolution. It is not true that all species that contain PI31 also contain Fbxo7 for example *Dictyostelium discoideum* contains a PI31 protein but an Fbxo7 protein has not been identified. If these two proteins evolved together to perform a function it would be expected that they would be found within the same species. It is possible that while PI31 and Fbxo7 may interact in the cell they may also have additional functions to perform alone.

The structure based sequence alignment of the FP domain appears to show that the conserved residues are not functionally active residues but they do appear to be required for the structural integrity of the protein. This requirement for these structurally important amino acids to be conserved points to the possibility that the overall structure of the FP domain is functionally required. It would also be interesting to examine the effect of mutating residues on the surface of the protein near the highly conserved Trp58 as it is possible that this could be part of a binding surface for Fbxo7 or for the proline rich C-terminus of PI31 itself.

The novel α/β fold of the PI31 FP domain does not belong to any previously identified family of proteins. It is interesting that it has a superficially similar arrangement to ubiquitin (β-grasp fold), which consists of a four stranded anti-parallel β-sheet with three α helices packed against one side (Figure 4.16). The PI31 FP domain consists of a five stranded anti-parallel β-sheet with five α helices packed against one side. However the topology of ubiquitin is very different to the PI31 FP domain as ubiquitin has a topology of ββαaβαβ and PI31 FP domain of ααββββββαaα and so this similarity does not point to any evolutionary relationship.

Figure 4.16 – Structure of ubiquitin

Ubiquitin shown in cartoon representation blue at N-terminus to red at the C-terminus. (PDB accession code 1OGW (Vijay-Kumar *et al.* 1987)).

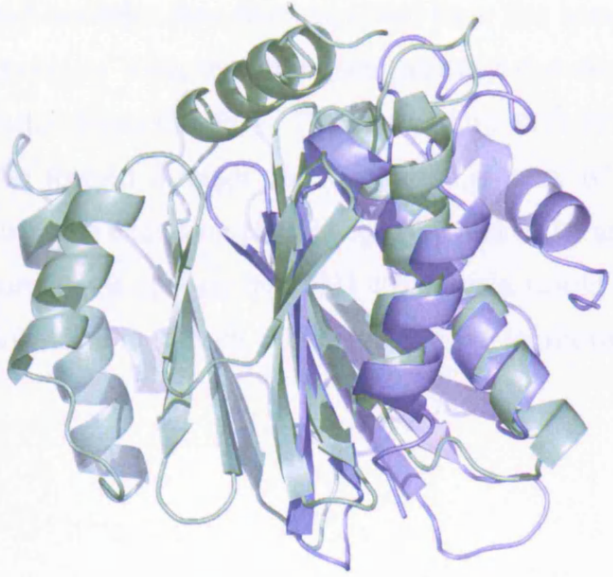

Figure 4.17 – Superposition of α subunit of *T.acidophilum* proteasomal subunit with PI31 FP domain

Alpha subunit of the *T.acidophilum* proteasome (green) (PDB accession code: 1PMA) superposed with PI31 FP domain monomer (blue). The 54 amino acids superposed have an RMSD of 3.14Å although the topology of the structures is not the same.

Interestingly PI31 FP domain also demonstrates superficial structural similarity with the 20S proteasomal subunits (Figure 4.17). The superposition of the α subunit of *T.acidophilum* with the PI31 FP domain reveals that 54 C-alpha residues superpose with an RMSD of 3.14Å. The topology of the 20S proteasomal subunits and PI31 is not the same and so there is unlikely to be an evolutionary link between the two proteins. This could be very interesting as the mode of action for the reported inhibition of the proteasome by PI31 has not been established. It is possible that the globular FP domain of PI31 could mimic the proteasomal subunit and bind within the proteasome. After binding the unstructured C-terminus of PI31, which was shown to be required for proteasomal inhibition, could be positioned inside the proteasome thus restricting access to the protealytically active β subunits.

The data from the protein-protein interaction servers indicated that neither interface A nor B were reliably within the values expected for a biologically relevant homodimer interface. The expected values for the homodimer interface were calculated using a sample set of 32 well defined homodimers (Jones and Thornton 1995). This is not a large number of proteins and so it is possible that the true values for homodimers may have a larger standard deviation than that calculated from this sample set. Interface A fits the expected parameters better than B, which indicated that this was more likely to form the correct solution dimer interface. The dimer of the PI31 FP domain was shown in this chapter to be formed through the N-terminal α helix of the two protomers (interface A) by using size exclusion chromatography and AUC analysis of both wild type and mutant proteins. It appears that PI31 FP domain would be likely to be in a monomer-dimer equilibrium in the cell, which would have implications for the function of this protein.

5 Characterisation of the PI31-Fbxo7-Skp1 interaction

The interaction between PI31 and Fbxo7 was established *in vitro* in chapter two and validated by co-immunoprecipitation and yeast two-hybrid. In this chapter the interaction between Fbxo7-Skp1 is characterised in more detail using biophysical and cell biology methods. A description of the experiments carried out is included as a flowchart in Figure 5.1.

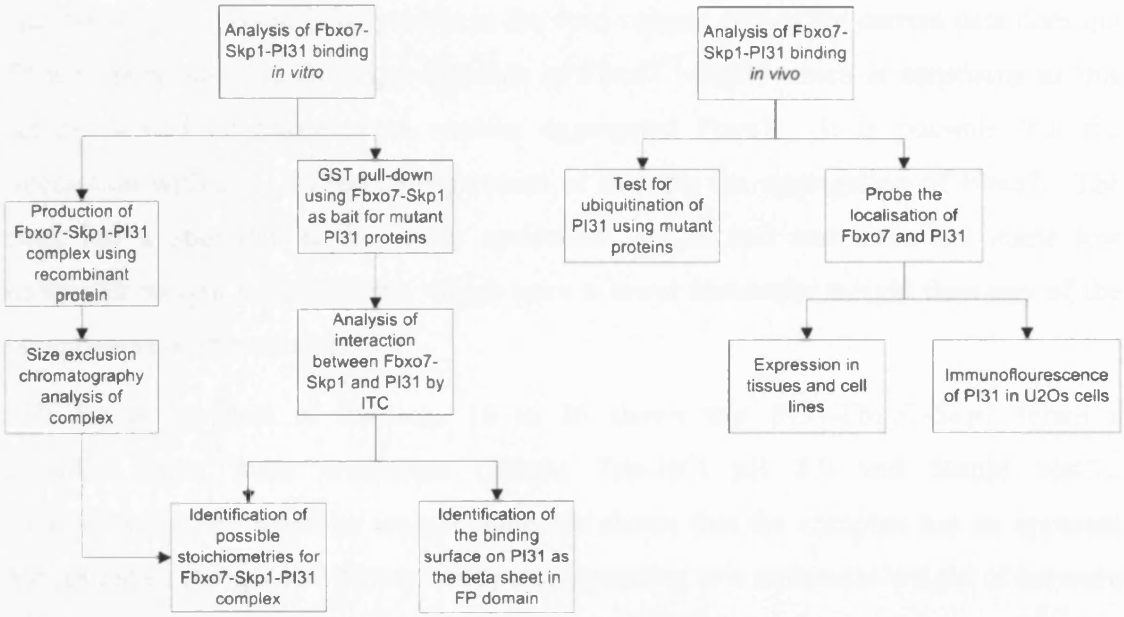

Figure 5.1 – Flowchart of the experiments discussed in chapter five.

5.1 PI31 mutations and characterisation of the interaction between PI31-Fbxo7-Skp1 *in vitro*

5.1.1 Formation of a PI31-Fbxo7-Skp1 complex

The interaction between Fbxo7 and PI31 was originally identified between recombinant, purified Fbxo7^{F2}-Skp1 and full-length endogenous PI31. The presence of the FP domain in both PI31 and Fbxo7 suggested that these domains may mediate the interaction between PI31 and Fbxo7. The Fbxo7^{F3}-Skp1 construct contains the FP domain connected to the F-box motif via a linker sequence whilst the PI31-NTD construct contains only the FP domain. These two constructs were used to produce recombinant proteins to further analyse the interaction(s) made between FP domains.

GST-tagged Fbxo7^{F3}-Skp1 protein complex was used as bait in a GST pull-down (appendix A) with recombinant PI31-NTD as a potential binding partner. The P131-NTD that co-purified with GST-Fbxo7^{F3}-Skp1 was visualised using SDS-PAGE and Coomassie brilliant blue staining. The three proteins formed a complex (Figure 5.2a). We exploited a thrombin protease site between the GST moiety and Fbxo7 to specifically elute the complex from the glutathione-Sepharose and GST label.

Analytical size exclusion chromatography was used to determine the stoichiometry of the components in the complex (Figure 5.2c). The complex elutes from the column as one major peak. There is no protein in the void volume and so the current data does not fit with previous data from gel filtration of Fbxo7^{F3}-Skp1, which is surprising as this sample would be expected to contain aggregated Fbxo7. It is possible that the interaction with PI31 FP domain prevents or disrupts the aggregation of Fbxo7. The peak has a shoulder at the lower molecular weight end and there are some low molecular weight contaminants, which have a lower molecular weight than any of the components of the complex.

SDS-PAGE analysis of fractions 16 to 26 shows that PI31-Fbxo7-Skp1 forms a complex under these conditions (20mM Tris-HCl pH 8.0 and 50mM NaCl). Comparison with molecular weight standards shows that the complex has an apparent Stokes radius (Siegel and Monty 1966) corresponding to a molecular weight of between 43kDa and 66kDa.

The shoulder on the eluted peak from fractions 24 to 26 consists of some Fbxo7^{F3} and also excess PI31 FP domain. This could be because excess PI31 FP domain was applied when the original GST pull-down was carried out or it could indicate that Fbxo7 and PI31 interact in the absence of Skp1. There is also an indication that Fbxo7 is undergoing degradation although this is surprising, as the limited proteolysis experiment described in section 2.1.2 did not show any further degradation to the domains used in this experiment.

Fractions 18-20 were pooled, concentrated to 5mg/ml and submitted to crystallisation trials; unfortunately no crystals were obtained.

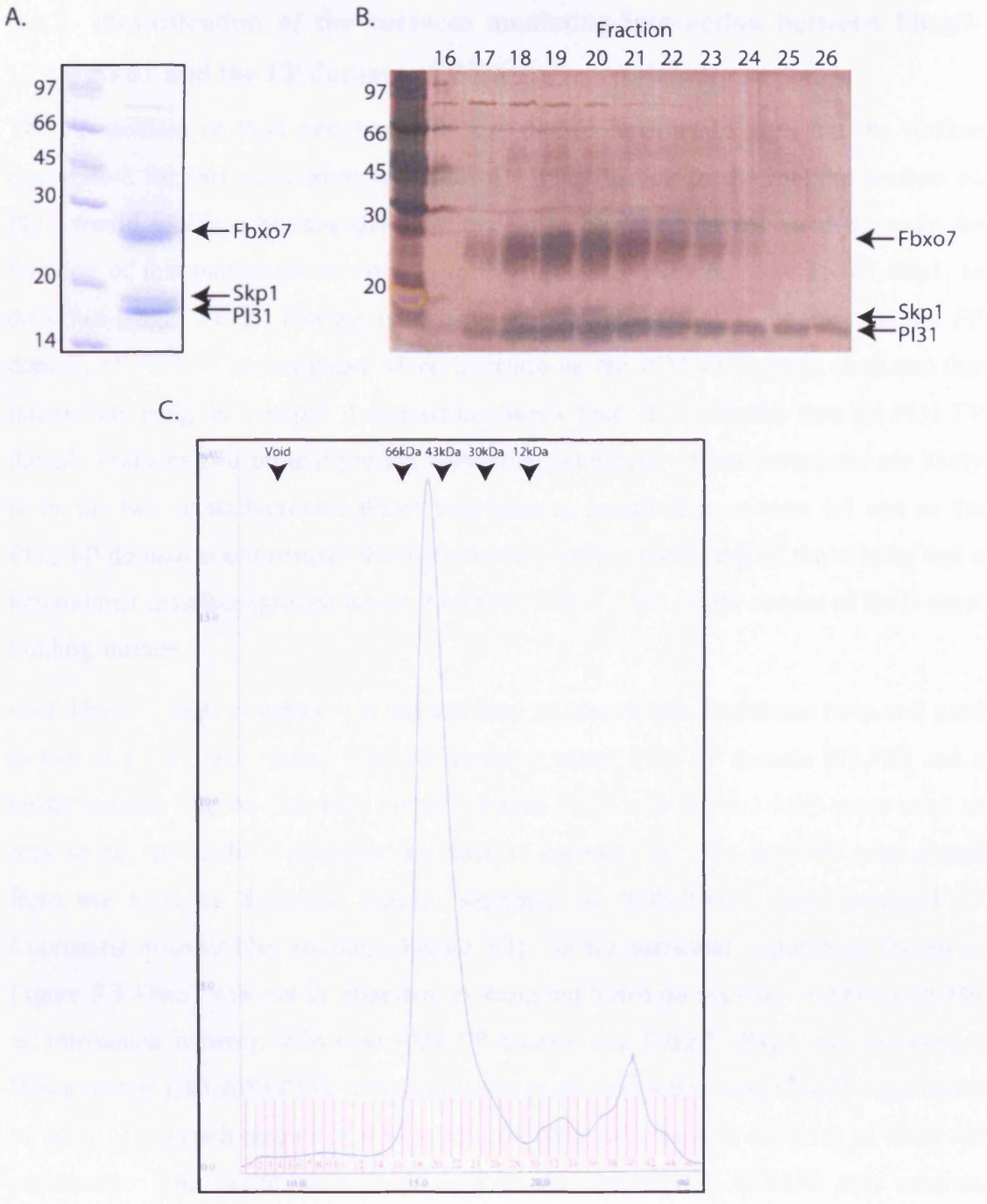

Figure 5.2 – Recombinant PI31 FP domain interacts with Fbxo7^{F3}-Skp1 protein complex.

A. GST pull down of PI31 FP domain using Fbxo7^{F3}-Skp1 as bait. Protein was eluted from the column by proteolytic cleavage. B. Silver stained SDS-PAGE of the fractions across the gel filtration profile shown in C. C. Size exclusion chromatography of PI31 FP domain–Fbxo7^{F3}–Skp1 complex. X-axis is volume in ml and y-axis is absorbance at 280nm measured in mAU. Gel filtration performed using a Superdex 75 HR 10/30 column (Amersham Biosciences) at a rate of 0.4 ml/min using 20mM Tris-HCl pH 8.0 and 50mM NaCl as buffer. 500µl of protein sample was loaded onto the column.

5.1.2 Identification of the surfaces mediating interaction between Fbxo7-Skp1 and the FP domain of P131

The FP domain of PI31 and Fbxo7^{F3}-Skp1 directly interact *in vitro* but the surface responsible for this interaction is unknown. Identification of the binding surface on PI31 would enable a binding deficient mutant to be produced and used to study the function of this interaction *in vivo*. The FP domain of PI31 binds to Fbxo7-Skp1, as does full-length P131. Having established Fbxo7-Skp1 and PI31 interact via the FP domain of PI31 we investigated which interface on the PI31 FP domain mediated this interaction using the mutants described in chapter four. It is possible that the PI31 FP domain contains two protein-protein interaction interfaces. These interfaces are likely to be the two crystallographic dimer interfaces as described in section 4.4 and so the PI31 FP domain could contain the homodimer interface consisting of the α-helix and a heterodimer interface through which it binds to Fbxo7-Skp1 might consist of the β-sheet binding surface.

GST-Fbxo7^{F3}-Skp1 complex was immobilised on glutathione Sepharose resin and used as bait in a GST pull down. PI31 FP domain, mutant PI31 FP domain I83,90E and a buffer control (20mM Tris-HCl pH 8.0, 50mM NaCl and 5mM 2-ME) were used as prey to the bait under conditions described in appendix A. The proteins were eluted from the resin as described before, separated by SDS-PAGE and visualised by Coomassie brilliant blue staining (Figure 5.3). In the particular experiment shown in Figure 5.3 Fbxo7 was not as abundant as expected based on previous experiments but an interaction between wild type PI31 FP domain and Fbxo7^{F3}-Skp1 was detectable. When mutant (I83,90E) P131 was used however, no interaction with Fbxo7-Skp1 could be seen. The stoichiometry of this interaction does not appear to be 1:1:1 as observed previously. This preliminary result suggests that interface B in P131 may mediate binding to Fbxo7-Skp1.

The GST pull down suggested that interface B on PI31 FP domain is the binding site for Fbxo7-Skp1. This experiment was not performed in physiological conditions and the presence of the glutathione Sepharose may have had an effect on binding. Isothermal titration calorimetry (ITC) is a thermodynamic technique for monitoring the heat energy release by the chemical process initiated by the addition of a binding partner, and this method requires no chemical tagging or immobilisation of the proteins (appendix A).

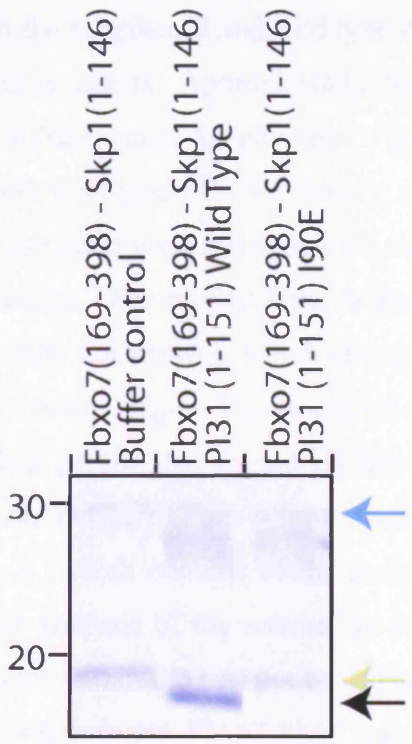

Figure 5.3 – SDS-PAGE of GST pull down to identify the Fbxo7 binding surface on PI31

Arrows mark the expected position of Fbxo7 (blue), Skp1 (green) and PI31 (black)

When proteins bind, heat is either generated or absorbed. Measurement of this heat allows accurate determination of binding constants (K_a), reaction stoichiometry (n), enthalpy (ΔH) and entropy (ΔS) providing a profile of the molecular interaction (Equation 5.1 and Equation 5.2). ITC is the most quantitative means available for measuring the thermodynamic properties of a protein-protein interaction because it measures the binding equilibrium directly (Wiseman *et al.* 1989; Pierce *et al.* 1999).

Equation 5.1

$$Q = \frac{V_0 \Delta H_b [M]_t K_a [L]}{1 + K_a [L]}$$

Where Q=heat evolved on addition of ligand, V_0=volume of the cell, ΔH_b=enthalpy of binding per molecule of ligand, $[M]_t$=total macromolecule concentration including bound and free fractions, K_a=binding constant and [L]=free ligand concentration.

Equation 5.2

$$c = K_a [M] n$$

The experimental binding isotherm can be characterised by the unitless value c, which is dependent on the stoichiometry of the reaction (n).

Fbxo7^{F3}-Skp1 was placed in the sample cell and wild type PI31 FP domain and I83,90E FP domain mutant were used as titrants. Approximately 20μM of Fbxo7^{F3}-Skp1 was in the cell and 100μM of PI31 FP domain in the injectant. Figure 5.4a and b show the raw data in the top panels for both wild type PI31 FP domain and I83,90E mutant PI31 FP domain. The raw data is on the same scale and shows the significant difference between the wild type and mutant proteins. The profile of the mutant protein is the same as that for the heating of dilution (data not shown), which strongly indicates that the mutant protein does not interact with Fbxo7-Skp1. The processed data for the PI31 FP domain I83,90E mutant has a random profile that cannot be fitted to any model. This data indicates that the interaction between Fbxo7-Skp1 requires that PI31 FP domain contains an intact interface B, which consists of the β-sheet of the PI31 FP domain. This data does not allow the analysis of the interaction interface on Fbxo7-Skp1 nor does it mean that other domains within PI31 do not contribute to this interaction. Figure 5.4a shows that the interaction between Fbxo7-Skp1 and wild type PI31 FP domain produces a detectable signal which can be fitted with a smooth best-fit curve when modelled with the assumption that there is one binding site. It is a requirement for ITC data that the reaction reaches saturation (determined by the flat line at the higher molar ratio) and if this is the case the gradient of the curve can be used to determine the stoichiometry and binding constants for the reaction. The stoichiometry of the PI31-Fbxo7^{F3}-Skp1 interaction as determined by ITC is probably 1:2 (PI31:Fbxo7-Skp1) as n=0.444±0.006 (Figure 5.4a) indicating that one PI31 protomer interacts with two Fbxo7-Skp1 heterodimers. This is unexpected because PI31 is in a monomer-dimer equilibrium in the cell (chapter 4) and so it would have seemed likely that one PI31 homodimer would interact with one Fbxo7-Skp1 heterodimer giving a stoichiometry of 2:1(PI31:Fbxo7-Skp1). The K_d for the binding of PI31 FP domain to Fbxo7^{F3}-Skp1 was determined by ITC as 0.37μM, which is a 10 fold stronger interaction than the homodimer formation of PI31 FP domain.

The PI31-Fbxo7-Skp1 complex formed during ITC analysis was purified using size exclusion chromatography to remove excess PI31 and concentrated to about 5mg/ml prior to being used in crystallisation trials. No crystals were obtained using this method.

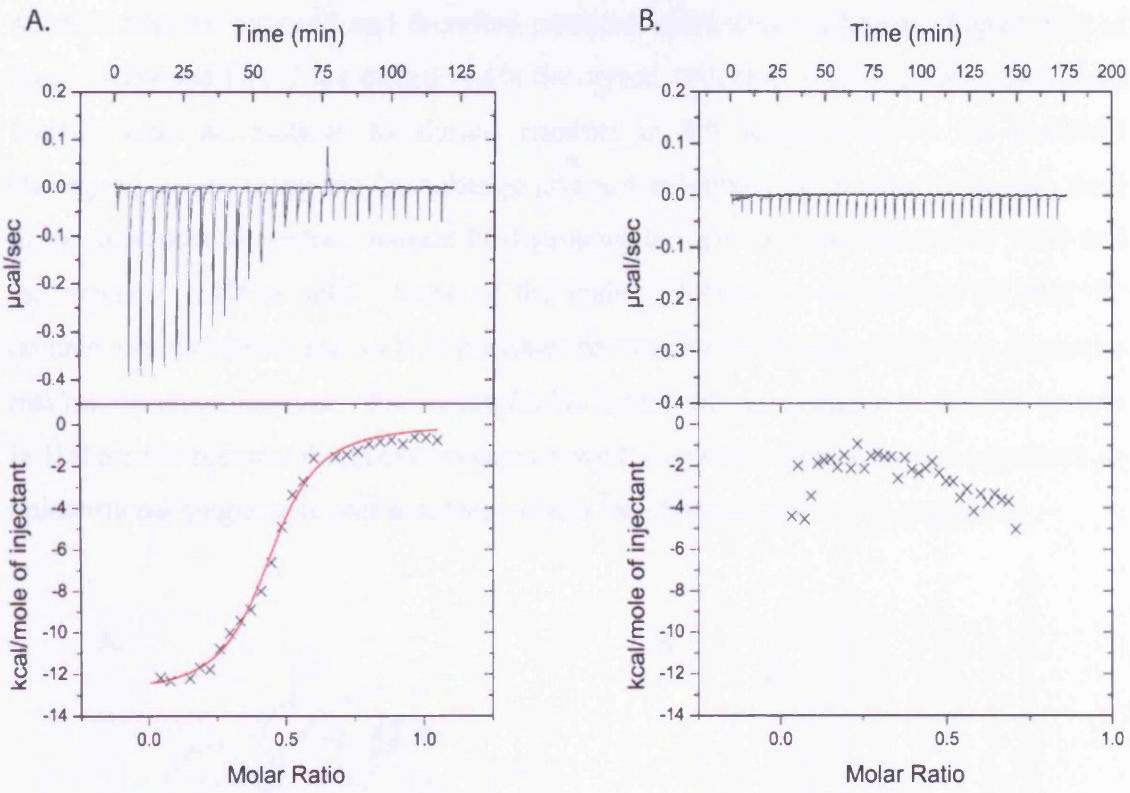

Figure 5.4 – ITC of Fbxo7^{F3}-Skp1 interaction with the PI31 FP domain wild type and I83,90E

Plot of the ITC data, the top panel is raw data after removal of the contribution by the heating of dilution. The bottom panel consists of the experimental points (x) and the solid red line corresponds to the best fit curve obtained by least-squares deconvolution. A) Wild type PI31 FP domain as titrant. B) I83,90E PI31 FP domain as titrant. The best fit values for the wild type PI31 FP domain are n=0.444±0.006 and K_d=0.37µM.

5.2 Characterisation of the interaction of PI31 with Fbxo7-Skp1 *in vivo*

5.2.1 PI31–Fbxo7-Skp1 interaction is required for ubiquitination of PI31

Fbxo7-Skp1 and PI31 have been shown to interact both *in vitro* and *in vivo* but the functional significance of this interaction has not been established. There are several potential functional roles for this interaction; PI31 is a substrate of an E3 ligase, PI31 is an E3 ligase co-factor, and PI31 is an E3 ligase localisation factor. In the former assumption the protein-protein interaction between an F-box protein and its substrate often catalyses the formation of an isopeptide bond between the substrate (or for polyubiquitination, ubiquitin) lysine residue and the C terminus of ubiquitin (Pickart 2001). Analysis of the structure of the PI31 FP domain revealed several possible lysine

residues that are exposed and therefore potential ubiquitination targets (Figure 5.5) of these Lys49 and Lys72 are disordered in the crystal structure. Lys49, Lys81, Lys72 and Lys125 were all mutated to alanine residues in full length PI31 in the pcDNA3 (Novagen) vector using the Quikchange protocol (appendix A) and Heike Laman used these constructs to produce mutant PI31 proteins to look for ubiquitination of PI31 and the mutants in HeLa cells. None of the lysine residues on the surface of PI31 FP domain are conserved and so it is possible that there is no requirement for a particular residue for ubiquitination. The construct for Lys49 did not produce detectable protein in HeLa cells because the primer sequence used to produce this construct contained an unintentional single base pair insertion, which introduced a stop in the sequence.

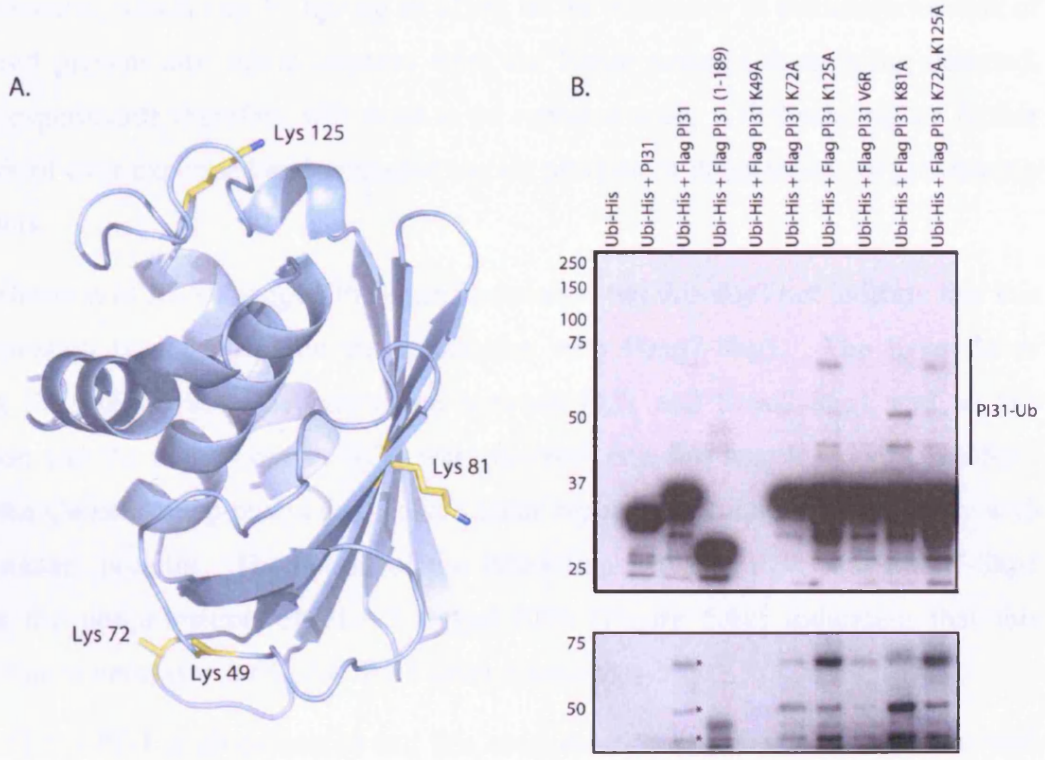

Figure 5.5 – Identification and testing of possible lysine residues for ubiquitination

A. Cartoon representation of the PI31 FP domain with all accessible lysine residues labelled. B. Over expression of His-tagged ubiquitin (Ubi-His) and PI31 in HeLa cells, His-tag pull down to detect ubiquitinated proteins and Western blot with the monoclonal anti-PI31 antibody to detect PI31. Panel below is over exposure of the section of the top panel marked with a line representing the ubiquitinated proteins. His-tag pull down and western blot performed by Heike Laman.

All mutant proteins except K49A are ubiquitinated to the same extent as wild type FLAG-tagged protein (Figure 5.5b). It is not possible to detect endogenous ubiquitinated PI31 in this assay (lane 1, Figure 5.5b), which is probably because the

quantity of protein is much lower than for the over expressed FLAG-tagged protein and no chemical proteasome inhibitor was included in this assay to allow the accumulation of ubiquitinated species including PI31. Surprisingly the over expression of untagged PI31 does not allow the ubiquitination of exogenous protein to be detected (lane 2, Figure 5.5b) although this may again be because the proteasome was not inhibited and the rapid turnover of polyubiquitinated PI31 may not allow its visualisation. To explain the apparent differences between untagged and FLAG tagged PI31 one hypothesis is that the FLAG tag itself is interfering with the ubiquitination. The FLAG tag contains two lysine residues and so the two FLAG tags in pcDNA3 (MDYKDDDDKDYKDDDDK) introduce an additional four lysine residues for ubiquitination, which may be having an effect on the efficiency of the ubiquitination of PI31 and prevent any subtle changes from the lysine mutants from being detected. These experiments therefore will need to be repeated using a different tag or further analysis of over expressed and untagged mutant proteins in the presence of proteasome inhibitors.

Ubiquitination of FLAG-tagged PI31 can be detected but this does not indicate that this ubiquitination is dependent on the interaction with Fbxo7-Skp1. The heterodimer mutant I83,90E disrupts the interaction between PI31 and Fbxo7-Skp1 and so this mutation and the single mutant I83E were inserted into full length PI31 in pcDNA3 using the Quikchange protocol and Heike Laman repeated the ubiquitination assay with these mutant proteins. Disruption of the interaction between PI31 and Fbxo7-Skp1 ablates the ubiquitination of FLAG-tagged PI31 (Figure 5.6a) indicating that this interaction is necessary for FLAG-PI31 ubiquitination.

Since FLAG-PI31 is ubiquitinated and this ubiquitination requires the interaction with Fbxo7-Skp1, PI31 could be directly or indirectly a substrate of the SCF^{Fbxo7}. When substrate proteins are polyubiquitinated through the Lys48 of ubiquitin one possible fate for the substrate is for it to be targeted for destruction by the 26S proteasome (Hershko and Ciechanover 1998) other possible fates are signalled by monoubiquitination or by polyubiquitination through the Lys63 of ubiquitin. siRNA was used by Heike Laman to knock-down the Fbxo7 RNA levels and therefore protein levels in HeLa cells and the corresponding amount of PI31 protein was analysed by Western blot (Figure 5.6b). The siRNA depleted but did not remove all Fbxo7 in the cells but no detectable change can be seen in PI31 protein levels. This result could be due to several reasons, Fbxo7 may

not be responsible for the ubiquitination of PI31 alternatively the interaction between PI31 and Fbxo7-Skp1 is of a modest affinity (with a $K_d = 0.37\mu M$) and so the remaining Fbxo7 could be sufficient to allow the steady state turnover of PI31. Also the PI31 ubiquitination may not target PI31 for destruction (Weissman 2001) but may serve a separate function. It may be helpful to repeat this experiment using the levels of the known substrate of Fbxo7, HURP, as a control to establish the efficiency of the siRNA.

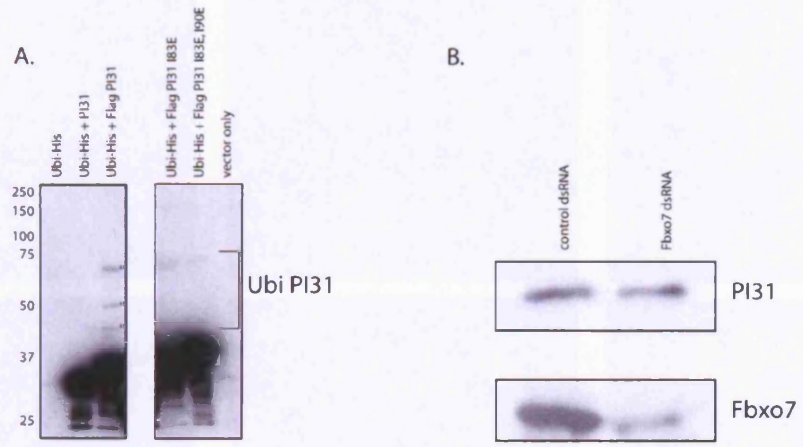

Figure 5.6 – PI31–Fbxo7-Skp1 interaction is required for PI31 ubiquitination but does not target PI31 to the proteasome

A. Experiment described in Figure 5.5a repeated with mutant PI31 protein. B. Protein content of HeLa cells with siRNA treatment using control dsRNA and Fbxo7 dsRNA analysed by Western blot with anti-Fbxo7 and anti-PI31. Figure provided by Heike Laman.

5.2.2 Localisation and expression of PI31 and Fbxo7

To better understand where or when these proteins interact *in vivo* an investigation of the cellular localisation of both the proteins was carried out. Over expressed untagged full length murine PI31 was shown by Zaiss *et al* (Zaiss *et al.* 2002) by immunofluorescence to be predominately localised at the cytosolic side of the nuclear envelope/endoplasmic reticulum (ER) in B8 mouse fibroblasts. Heike Laman repeated this experiment using endogenous PI31 in U2OS cells (Figure 5.7) and the localisation is again predominately at the ER although additional staining can also be seen. When Fbxo7 was over expressed (Laman *et al.* 2005) by transient transfection, the exogenous protein was localised in the cytoplasm, with small amounts being detected in the nucleus. Endogenous Skp1 was shown by Freed *et al* (Freed *et al.* 1999) to localise at the centrosome and in the nucleus and cytoplasm.

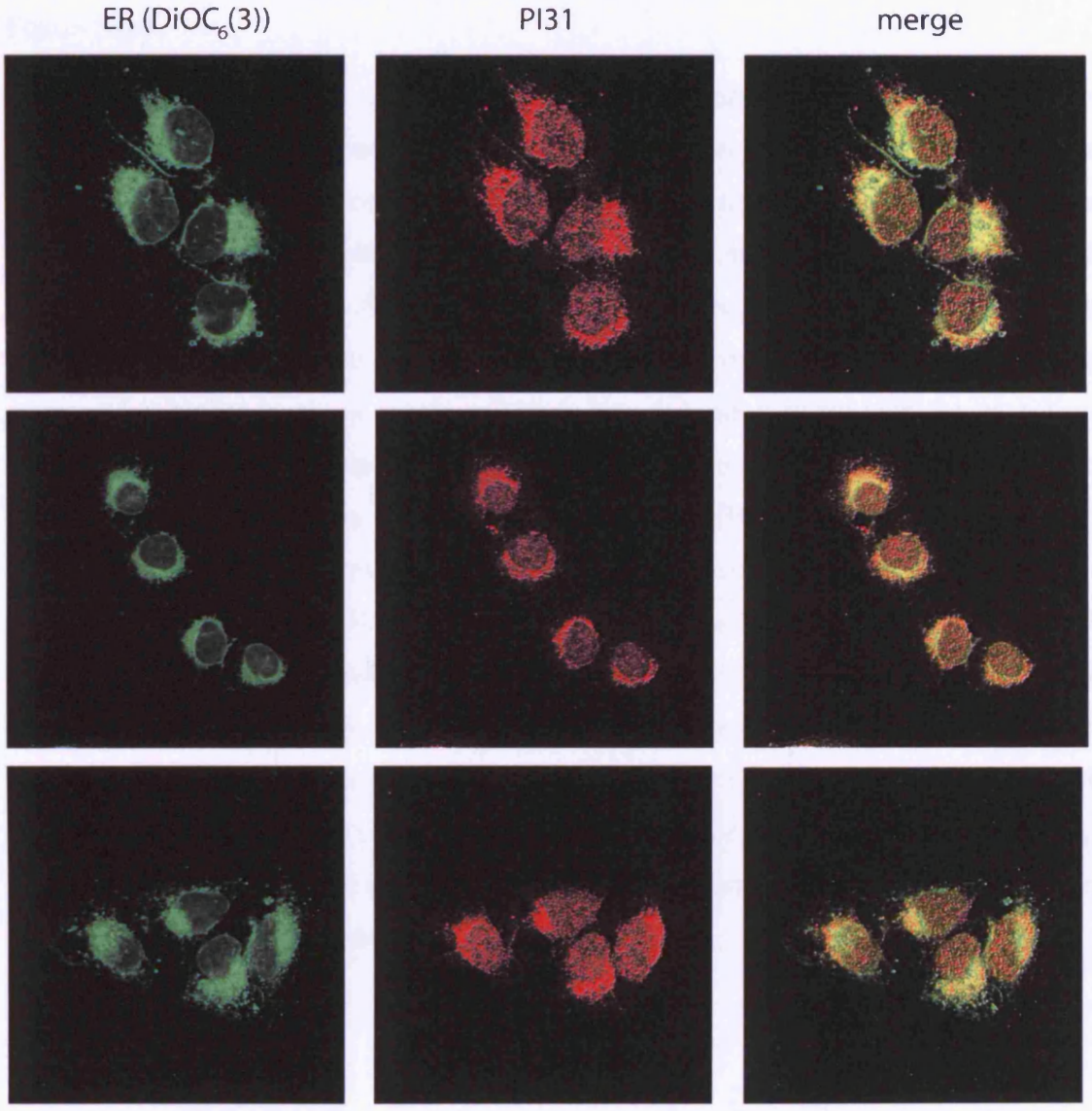

Figure 5.7 – PI31 co-localises with the ER

The ER was visualised in fixed U2OS cells by immunofluorescence staining with $DiOC_6(3)$ (green) and endogenous PI31 was visualised with affinity-purified monoclonal mouse anti-PI31 antibody and donkey anti-mouse rhodamine conjugated antibody (red). Right panel shows ER and PI31 superposed and co-localisation appears in yellow. Immunofluorescence work done by Heike Laman.

The localisation of the proteins indicates that the majority of Fbxo7-Skp1 would be unlikely to come into contact with PI31 in the cell. During mitosis in higher eukaryotes the nuclear envelope breaks down and the nuclear contents are completely mixed with the cytosol (Alberts *et al.* 1989), which would temporarily bring PI31 into contact with Fbxo7-Skp1.

If PI31 and Fbxo7 interact within the cell it can reasonably be expected that the two proteins would have the same or similar tissue distributions within the body. Western blot analysis of tissues and cell lines revealed that PI31 and Fbxo7 are expressed in the same tissues at similar levels other than in the placenta and ovary where only Fbxo7 was detected (Figure 5.8). Samples from healthy tissues were compared with cancer cell lines by Heike Laman and revealed that both Fbxo7 and PI31 appear to be expressed at higher levels in cancer cell lines than in healthy tissue with the exception of the proliferating HOB(6p) cell line, which is derived from normal human bone marrow. This difference in expression can not be confirmed as there is no loading control present on this Western blot and so this should be considered to be a preliminary result of the expression profiles of both Fbxo7 and PI31 and further work will be needed to characterise this using matched tissue samples. Fbxo7 was reported by Laman *et al* (Laman *et al.* 2005) to be an oncogene and so expression of Fbxo7 could be expected in cancer cell lines but there is no phenotype associated with over or under expression of PI31. Both Fbxo7 and PI31 are expressed in brain tissue, which is non-proliferating, and so more work is needed to compare the levels in tumour and non-tumour tissues to draw conclusions from this data.

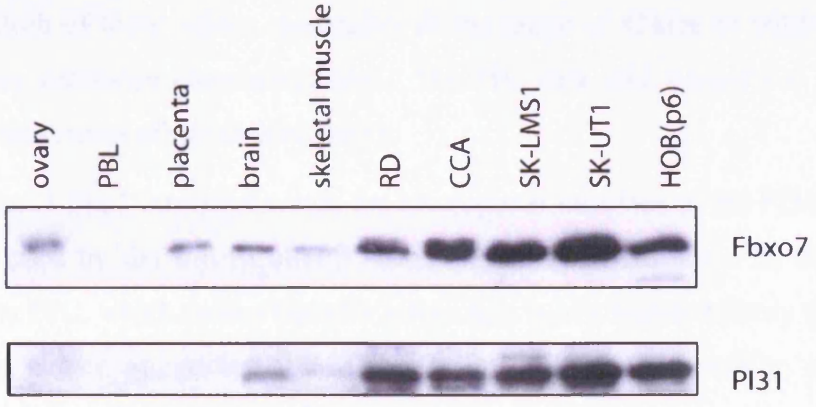

Figure 5.8 – Expression of PI31 and Fbxo7 in various tissues and cell lines

Expression of PI31 and Fbxo7 in tissues and cancer cell lines. Proteins detected by Western blot using anti-Fbxo7 and anti-PI31 by Heike Laman.

5.3 Discussion

The interaction between Fbxo7-Skp1 and the FP domain of PI31 has been shown both *in vivo* and *in vitro*. It remains to be determined if the proline rich C-terminus of PI31 contributes to the binding. The FP domain is sufficient for binding to Fbxo7-Skp1 but the affinity for full length PI31 protein has not been shown here. It is also possible that *in vivo* this interaction is modulated by other proteins that have yet to be identified.

Fbxo7^{F3}, Skp1ΔH8 and PI31 FP domain proteins have predicted molecular weights of 24.9kDa, 16.5kDa, and 16.9kDa respectively. The stoichiometry of the three proteins in a ternary complex as analysed by size exclusion chromatography does not appear to be 1:1:1 as there is less Skp1 than either Fbxo7 or PI31. However the Fbxo7^{F3}-Skp1 protein complex used in this interaction was not purified by size exclusion chromatography prior to exposure to PI31 and so it is possible that there is aggregated Fbxo7 in the sample which can still bind PI31 in the absence of Skp1. If a PI31 FP domain:Fbxo7^{F3}:Skp1 complex existed as a 1:1:1 stoichiometric complex it would be predicted to have a molecular weight of 58.3kDa which correlates well with the experimentally determined size of about 55kDa. A 2:1:1 complex (75.2kDa), 1:2:1 complex (83.2kDa) and a 1:1:2 complex (74.8kDa) do not fit with the experimental data. Although the size exclusion chromatography data suggest that a 1:1:1 complex forms *in vitro*, it is entirely possible that other stoichiometries exist *in vivo*. Another possibility raised by the excess of PI31 FP domain and Fbxo7 in the SDS-PAGE of the fractions is that the complex exists as PI31 FP domain:Fbxo7 alone. A 2:1 complex would be predicted to have a molecular weight of 58.7kDa and a 1:2 complex of 66.7kDa. Both of these values reasonably fit the range of 43kDa to 66kDa calculated from the size exclusion chromatography. The ITC data also supports a 1:2 complex from its measurement of the stoichiometry.

PI31 and Fbxo7-Skp1 interact through the heterodimer interface of the PI31 FP domain, which is formed by the anti-parallel β sheet. This interaction has a K_d of 0.37μM as calculated by ITC, which means that this interaction is of a higher affinity than the PI31 FP domain dimer suggesting that Fbxo7-Skp1 could successfully compete for monomeric PI31. This may provide a better explanation whereby two Fbxo7-Skp1 heterodimer complexes bind to one PI31 FP domain monomer as suggested by ITC.

Preliminary data shows that FLAG-PI31 is ubiquitinated and that this ubiquitination requires Fbxo7-Skp1 binding through analysis of PI31 mutant defective in binding to Fbxo7-Skp1. Depletion of Fbxo7 has no noticeable effect on PI31 protein levels. The ubiquitinated lysine residue(s) on PI31 have not been identified which could be because the FLAG tag interfered with the interaction or it could be that more than one lysine residue on PI31 can be ubiquitinated as has been observed in other proteins (Hod *et al.* 1994).

PI31 and Fbxo7 have not been shown to co-localise. A crucial aspect of validity of *in vivo* interactions is to show where and when the proteins interact. At present we do not understand whether PI31 and Fbxo7 can interact, one possibility is that the Ubl of Fbxo7 brings it into contact with the immunoproteasome where PI31 may be localised, or modified PI31 may be localised to the cytoplasm for Fbxo7 mediated degradation. It is of primary importance to address this problem with future work.

6 Discussion

6.1 Summary

One of the aims of this thesis was to structurally characterise the Fbxo7-Skp1 complex and to identify potential substrates of SCF^{Fbxo7}. Unfortunately it did not prove possible within the time of this thesis to obtain suitable diffracting crystals for structure determination. However biochemical work in parallel with the structural studies lead to the identification of PI31 as a putative substrate for SCF^{Fbxo7}.

The identification of a shared FP domain within both Fbxo7 and PI31 refocused our efforts on the structural characterisation of the FP domain of PI31. After overcoming problems encountered with obtaining experimental phases the structure of a L7M FP domain mutant from PI31 was solved by MAD phasing to a resolution of 2.64Å. The structure of the PI31 FP domain was compared to previously solved structures and shown to have a novel topology and to form dimers in solution predominately through the N-terminal α helix.

Size exclusion chromatography and ITC were used to show that PI31 and Fbxo7-Skp1 interact through a surface on the anti parallel β-sheet of the PI31 FP domain. It has been inferred from a second crystallographic dimer that the same interface within the Fbxo7 FP domain mediates this interaction with PI31 although this remains to be proven. We have not yet shown where or when PI31 and Fbxo7-Skp1 might co-localise in the cell and so it has not yet been possible to prove a functional role for this interaction; ongoing studies will address these concerns. Ubiquitination of P131 is dependent on its interaction with Fbxo7-Skp1 but depletion of Fbxo7 has apparently no noticeable effect on PI31 protein levels.

There are several questions about this interaction that remain to be addressed with future work:

1. When and where do PI31 and Fbxo7 co-localise and does localisation play a role in the function of these proteins?
2. Does Fbxo7 directly or indirectly ubiquitinate PI31 to signal its destruction in the cell?
3. What is the stoichiometry of the interaction between Fbxo7 and PI31?

4. Is there a dominant-negative phenotype for over expression of PI31 mutant *in vivo*?
5. In the longer term does PI31 or Fbxo7 inhibit the maturation or formation of the immunoproteasome and is the β or α interface of the PI31 FP domain required for immunoproteasome formation?
6. How is SCF^{Fbxo7} regulated?

6.2 Stoichiometry of the PI31-Fbxo7-Skp1 interaction

Results from gel filtration and ITC experiments revealed potentially conflicting data about the stoichiometry of the interaction between PI31 and Fbxo7-Skp1. It appears from the ITC data that one molecule of PI31 binds to two molecules of Fbxo7-Skp1. Since PI31 is in a monomer-dimer equilibrium in cells, as established by analytical ultracentrifugation, this stoichiometry is a little surprising. It is possible that a monomeric form of PI31 binds to two Fbxo7-Skp1 heterodimers or that four Fbxo7-Skp1 complexes bind to one PI31 dimer. Fbxo7-Skp1 complex alone appears by gel filtration to have a 1:1 Fbxo7:Skp1 stoichiometry, that is a single heterodimer of 43kDa. This implies that either there are two Fbxo7-Skp1 binding sites on each PI31 protomer or that PI31 induces oligomerisation of Fbxo7-Skp1. Disrupting the β-sheet surface of PI31 by isoleucine to glutamic acid mutations completely abolishes binding to Fbxo7-Skp1 and so if there are two binding surfaces for Fbxo7-Skp1 on PI31 (Figure 6.1) then there may be a requirement for binding through this β-sheet interface first in order to allow the second molecule to bind; i.e. binding may be cooperative.

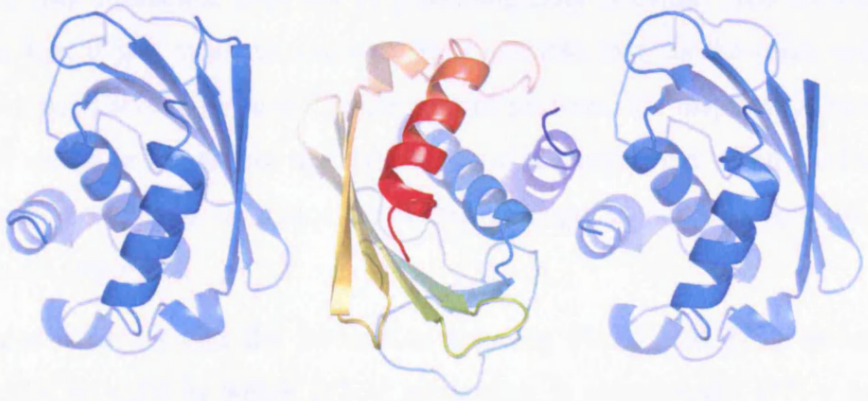

Figure 6.1 – Possible model for the interaction between the FP domains of Fbxo7 and PI31

Model using the FP domain of PI31 to demonstrate one possible arrangement of PI31 FP domain (centre) interacting directly with two molecules of Fbxo7 through their FP domains.

The monomeric form of PI31 could be used to test this model as this mutant protein disrupts the α helix homodimeric binding interface. If this interface is involved in binding to a second Fbxo7 FP domain analysis by ITC would show different stoichiometry and affinity constants. Gel filtration or analytical ultracentrifugation experiments could also be used to provide an estimated molecular weight of the complex using both wild type and mutant proteins and the stoichiometry calculated from this.

Structural information of the PI31-Fbxo7-Skp1 complex would address the seeming contradictions between the ITC and size exclusion chromatography data for the stoichiometry of this interaction as well as showing the molecular details of this interaction.

It is tempting to infer that the interaction is driven by the association between the FP domains of PI31 and Fbxo7 but we do not have experimental evidence for this at this stage. *In vivo* experiments will also be needed to show if the interaction between Fbxo7-Skp1 and PI31 is through Fbxo7 alone, Skp1 alone or the heterodimeric complex. This could be performed by co-immunoprecipitation using an Fbxo7 construct in which the F-box domain is deleted; this would then demonstrate if the interaction could proceed in the absence of Skp1 binding to Fbxo7.

6.3 Localisation of PI31

The mode of interaction between Fbxo7-Skp1 and PI31 is potentially very interesting for studies performed *in vitro* but if the proteins do not co-localise in the cell the function of this interaction may not be physiologically relevant. We are pursuing the hypothesis that it is a valid interaction. It is possible that, as the other substrates of Fbxo7 have been shown to be cell cycle control proteins, the interaction between PI31 and Fbxo7 could be dependent upon the stage of the cell cycle. Immunofluorescence studies on the proteins on cells blocked at various stages in the cell cycle would help to show if this is important.

Another possibility is that the interaction between Fbxo7 and PI31 occurs only in infected cells or cells in which IFN-γ production is stimulated. IFN-γ induces the formation and maturation of the immunoproteasome which is inhibited or blocked by PI31, therefore it is possible that ubiquitination of PI31 by SCF^{Fbxo7} is triggered by viral attack and the release of this cytokine. Using the mutant PI31 proteins and analysing

MHC class I presented peptides produced only by the immunoproteasome might be used to test this possibility.

The majority of Fbxo7 has been shown to be in the cytoplasm and the majority of PI31 in the ER however it is possible that a small but significant amount could be co-localised and the signal masked by the higher signal from other areas in the cell. Density gradient centrifugation and Western blotting for Skp1, Fbxo7, PI31 and carefully chosen marker proteins to identify the cellular compartments could reveal if this possibility is true.

6.4 Ubiquitination and expression of PI31

Although the interaction between PI31 and Fbxo7 is required for the ubiquitination of FLAG-PI31 the reduction of Fbxo7 by siRNA does not affect the levels of PI31 protein in the cell. One explanation for this result could be that Fbxo7 normally targets only a small fraction of the total pool of PI31 for degradation and so the global protein levels do not change significantly when Fbxo7 is reduced. Another possibility is that if PI31 is ubiquitinated and not subsequently destroyed then the ubiquitin chain may be formed through the Lys63 of ubiquitin and so it is not recognised for destruction by the proteasome. This can be tested *in vivo* using ubiquitin with a K48A or K48R mutation which are unable to form the Lys48-linked polyubiquitin chains but are still able to form the Lys63-linked polyubiquitin chains as well as monoubiquitination on several lysine residues. These post translational modifications often have alternative fates for the substrate protein. *In vitro* ubiquitination experiments using recombinant proteins could also reveal if SCF^{Fbxo7} is sufficient for the ubiquitination of PI31. This is not a trivial experiment as the relevant E2 enzyme must be identified first and the recombinant proteins of Fbxo7-Skp1 are prone to aggregation.

If the lysine targeted for ubiquitination on PI31 can be identified it would be possible to analyse the phenotype that a stable PI31 protein would produce. It is possible that the ubiquitination of PI31 is specific for one lysine but it has been reported that ubiquitination can occur on more than one lysine for some proteins (Hod *et al.* 1994). FLAG tags contain four lysine residues, which could be ubiquitination targets that are masking the effects of the mutations. They are also tagging the N-terminus of the protein, which is the homodimer interface and could affect its interaction with other proteins. It would be interesting to repeat the analysis of the lysine to alanine mutants

in untagged PI31; however we have not been able to detect at present if untagged PI31 is ubiquitinated.

Fbxo7 has been reported to be the product of a proto-oncogene and so is over expressed in cancer cells (Laman *et al.* 2005). It would be interesting to look at the expression of both PI31 and Fbxo7 in cancer and normal tissues and to see if the expression profiles of the two proteins are the same.

6.5 Analysis of the proline rich region of PI31

The proline rich region of PI31 was shown to be responsible for the inhibition of the 20S proteasome *in vitro* (Zaiss *et al.* 1999; McCutchen-Maloney *et al.* 2000). I have identified a motif in the C-terminus of PI31 (residues 232-254 in human) that is conserved and can be found in an apparently unrelated protein in *S. pombe* (Figure 6.2). Neither a PI31 nor a Fbxo7 protein has not been found in *S. pombe* and there are also no proteins identified as containing the FP domain. This proline rich motif may have a functional role indicating that PI31 may have a function in lower eukaryotes that do not contain the immunoproteasome. The identification of this motif raises the possibility that the original function of PI31 is dependent on the C-terminus of the protein and that the function of the globular N-terminal portion of the protein has developed more recently in evolution.

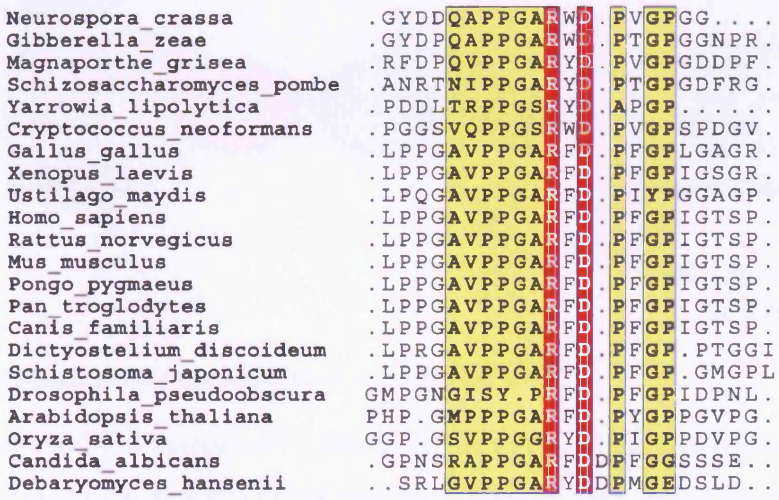

Figure 6.2 – Alignment of the C-terminal proline rich motif identified in PI31 and yeast proteins.

Protein sequences obtained by searching PSI-BLAST (Altschul *et al.* 1997) with human PI31 and aligned with ClustalX (Thompson *et al.* 1997). Final figure produced using ESPript (Gouet *et al.* 1999).

The GYF domain of the protein CD2BP2 was recently shown to bind the motif PPGxR in the proline rich C-terminus of PI31 (Kofler *et al.* In Press). The motif that Kofler *et al* identified is within the conserved proline rich motif identified here. The implication of this has yet to be elucidated but the identification of the GYF binding motif in lower eukaryotes may indicate an important functional role for this interaction.

To follow up these initial observations we have initiated a collaboration with the Cancer Research UK Cell Regulation Laboratory to explore the potential functional homologue of PI31 in *S. pombe* by over expression, deletion and tagging of this protein with GFP. It is hoped that this will identify a phenotype for the deletion or over expression of this protein and help to establish its localisation.

6.6 Model of the interaction between PI31 and SCF^{Fbxo7}

In spite of the questions that remain it is possible to use the current data to produce a plausible hypothetical model for the functional interaction between PI31 and Fbxo7 (Figure 6.3). From this model it is possible to make specific predictions that can be tested.

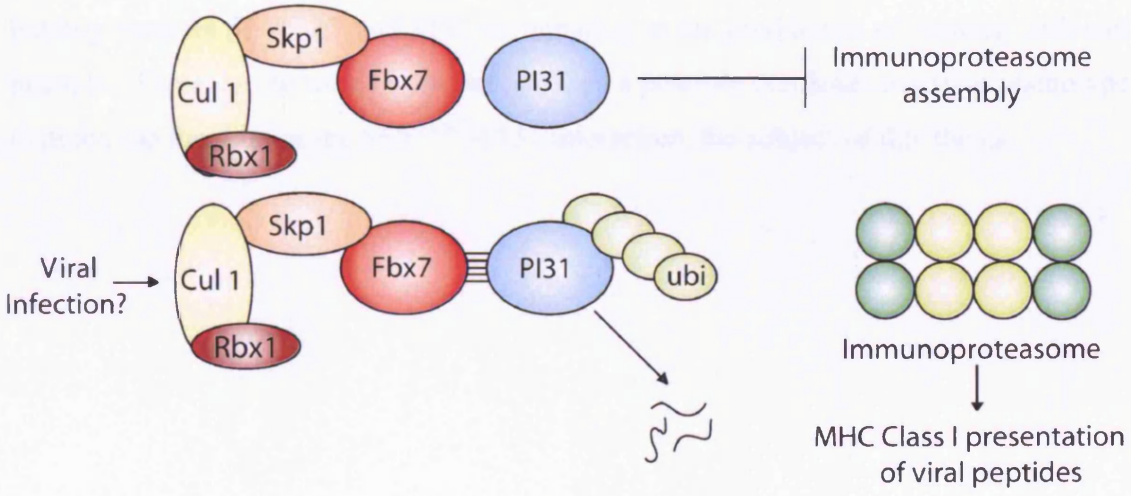

Figure 6.3 – Model of the interaction between PI31 and SCF^{Fbxo7}

As part of the model it conceivable that in the absence of viral infection PI31 is not targeted for degradation and it is likely that there is no interaction between SCF^{Fbxo7} and PI31. Under these conditions Fbxo7 is localised in the cytoplasm and is functioning to maintain the protein levels of HURP and Cdk6. PI31 is localised in the ER and inhibits the formation of the immunoproteasome in mammals.

Under the conditions of viral infection production of IFN-γ is induced and it acts as a signal to indirectly promote the interaction between PI31 and SCF^{Fbxo7}. This interaction could arise through recruitment of SCF^{Fbxo7} to the proteasome site where PI31 is localised. The Ubl domain of Fbxo7 may bring the SCF^{Fbxo7} into contact with the proteasome as the S5a subunit of the 26S proteasome contains a UIM that can bind to Ubl domains. The degradation of PI31 would remove its inhibitory role in the maturation of the immunoproteasome and so the immunoproteasome would assemble. The formation of the immunoproteasome enables the efficient production of non-self MHC class I peptides which are then processed and presented on the surface of the cell for detection by T cells.

The tight regulation of PI31 by SCF^{Fbxo7} may not be readily observed without further work analysing the degradation of PI31 using detailed biochemical and cell biology techniques. Some of the required experiments have been outlined in this chapter.

This thesis has shown a synergy of different techniques including structure determination, cell biology, bioinformatics and biophysical analysis. A considerable amount has been learnt concerning the domain structure, evolution and potential binding partners of Fbxo7 and PI31 culminating in the production of binding deficient mutants. These can be used *in vivo* to establish a possible dominant-negative phenotype to probe the function of the SCF^{Fbxo7}-PI31 interaction, the subject of this thesis.

Appendix A Materials and methods

A.1 Reagents and enzymes

Reagents were purchased from Merck or Sigma unless otherwise stated and were of the highest grade available. Enzymes used in molecular biology protocols were purchased from New England Biolabs (NEB), unless specifically stated.

A.2 Buffers, solutions and media

The common buffers and media used are described here; when a more specialised one was used it is described in the appropriate section.

Name	Composition
SDS sample buffer	2% (w/v) SDS, 80mM Tris-HCl pH 6.8, 10% (v/v) glycerol, 0.02% (w/v) bromophenol blue, 50µl/ml 2-ME added just before use
SDS-PAGE running buffer (10x)	15.1g Tris base, 94g glycine, 50ml 10% SDS made up to 1l with Milli Q grade reverse osmosis water
TAE buffer	40mM Tris-acetate, 2mM EDTA
TBE buffer	90mM Tris-borate, 2mM EDTA
LB medium	5g Bacto yeast extract, 10g Bacto-tryptone, 10g NaCl made up to 1l with sterile water and autoclaved.
TB medium	10g Bacto-tryptone, 5g NaCl made up to 1l with sterile water and autoclaved
NZY broth	5g NaCl, 10mM $MgCl_2$, 10g NZ amine, 5g Bacto yeast extract made up to 1l with sterile water and autoclaved.
L-AGAR	6g agar added to 400ml LB. Autoclaved for 15 minutes.

M9 media 10x	60g Na_2HPO_4 monohydrate, 30g KH_2PO_4 trihydrate, 10g NH_4Cl, 5g NaCl, 30mg $CaCl_2.2H_2O$ made up to 1l with sterile water and autoclaved.
SOC media	20g Bacto-tryptone, 5g Bacto yeast extract, 10mM NaCl, 2.5mM KCl, 10mM $MgCl_2$, 10mM $MgSO_4$, 20mM D-glucose made up to 1l with sterile water and autoclaved.
DNA sample buffer (6x)	0.25% bromophenol blue, 0.25% xylene cynaol FF, 30% glycerol.
Potassium salts	16.5mM K_2HPO_4 and 3.9mM KH_2PO_4 in sterile water and autoclaved.
Casamino acids	0.25g casamino acids (Gibco) made up to 1l with sterile water and 0.22μM filtered. Used at 4ml per l of NZY medium.

A.3 Cloning and PCR

A.3.1 Donated constructs

The following constructs were generously provided for this work. Human cDNAs encoding Skp1 and Fbxo7 were consecutively cloned into the expression plasmid pGEX-KG (Guan and Dixon 1991) to yield di-cistronic constructs which were provided by Heike Laman (Wolfson Institute for Biomedical Research, University College London, London , UK). This expression plasmid was adapted by Phillip Knowles (Structural Biology Laboratory, Cancer Research UK) to contain a TEV cleavage site. Human cDNAs of Cul1 (14-312) were cloned into pGex(KG) using the Nco1 and Xho1 restriction sites and this construct was also supplied by Heike Laman. Human cDNAs of PI31 (1-151) were cloned into the expression plasmid pET-28a(+) (Novagen) by Neil McDonald (Structural Biology Laboratory, Cancer Research UK).

A.3.2 Vectors

PGEX-KG was used to express all Fbxo7-Skp1 and Cul1 constructs; this system is discussed in section 2.1.1. The pET expression system (Novagen) was utilised for expression of PI31. The pET system was originally developed by Studier and Moffatt

(Studier and Moffatt 1986) and then commercially developed by Novagen. The expression plasmid pET-28a(+) contains an amino-terminal poly-histidine (His) tag for affinity purification, a thrombin cleavage site for removal of the His-tag and a kanamycin antibiotic resistance marker. It also contains the T7*lac* promoter, which consists of a *lac* operator sequence downstream of the T7 promoter as well as the promoter and coding sequence for the *lac* repressor (*lacI*), oriented so that the T7*lac* and *lacI* promoters diverge. FB810 cells (section A.4.1) contain a chromosomal copy of the gene for T7 RNA polymerase, the *lacI* gene and the *lacUV5* promoter. When the pET28 vector is used in FB810 cells the *lac* repressor acts at both the *lacUV5* promoter in the host chromosome to prevent the transcription of the T7 RNA polymerase gene and at the T7*lac* promoter to block the transcription of the target gene. IPTG can be used to induce the expression of T7 RNA polymerase under the control of the *lacUV5* promoter. The T7 RNA polymerase in turn transcribes the target DNA in the expression plasmid. This provides a tightly controlled expression system whereby the concentration of IPTG can be adjusted to optimise the expression of soluble target protein.

A.3.3 Oligonucleotides

Oligonucleotides were synthesised by the Cancer Research UK Oligonucleotide Synthesis Department and purified by ethanol precipitation. Oligonucleotides longer than 45 nucleotides were purified by HPLC. After closure of this facility in July 2004 all oligonucleotides were purchased from Sigma-Genosys.

A.3.4 Plasmid purification

Plasmid DNA was purified using either a Qiagen maxiprep kit or miniprep kit according to the manufacturers instructions. Plasmid concentration was calculated by measuring the absorbance at 280nm in a Pharmacia Biotech Ultrospec2000 spectrometer compared to a blank. Plasmid DNA was sequenced by PCR using the BigDye™ Terminator Cycle Sequencing kit (PE Applied Biosystems) according to the manufacturers instructions. Sequencing reactions were analysed by the Cancer Research UK equipment park by capillary sequencing using an ABI Prism 3730.

A.3.5 Polymerase Chain Reaction (PCR)

PCR amplification (Old and Primrose 1994) was carried out using an initial cycle of denaturation at 95°C for 5 mins, annealing at 3°C below the T_m of the primers for 4 mins and extension at 72°C for 4 mins. This was followed by 25 cycles of denaturation, annealing followed by extension for one minute at each of the three steps. Reactions typically contained 0.1μg template, 10pmole of each primer and 2 units of Vent DNA polymerase in a total volume of 50μl 1 x Thermopol buffer (NEB) supplemented with 1.5mM $MgSO_4$ and 0.5mM each of dATP, dCTP, dGTP and dTTP. If this procedure did not succeed in producing a clean DNA product the Optiprime kit (Stratagene) was used to optimise the PCR conditions.

Production of all single base pair mutations used in this thesis was performed using the Quikchange kit (Stratagene) and primers in Figure A.1

I90E	gtg gag agc agc atg gag ctc aat gtg ctg g
I83E	tc ctt gtg aaa gcc gag acc gtg gag agc
V6R	atg gcg ggc ctg gag cga ctg ttc gca tcg
L7M	g gcg ggc ctg gag gta atg ttc gca tcg gca gcg ccg g
K49A	ggc ggg tcc caa tga tgc gaa gtc aga act gct g
K72A	cct ccg gta tga gta tgc gga tgg gtc cag aaa g
K81A	cag aaa gct cct tgt ggc agc cat cac cgt gga ga
K125A	ctt cca cag gac cta cgc gaa cag tga gga gct t

Figure A.1 – 5' Primers used to create PI31 mutant proteins

3' primers are the reverse complement of the 5' primer in each case.

A.3.6 Restriction Digests

Restriction digests were performed using the supplied buffers at recommended temperatures (NEB). Typically, 1-2 units of enzyme were used per μg of DNA.

A.3.7 Agarose gel electrophoresis

DNA was analysed by electrophoresis on 0.8% agarose gels with 0.5μg/ml EtBr added and with TAE buffer if extraction of DNA was required or TBE buffer for a purely analytical gel. DNA was purified from agarose gels using the Ultrafree-DA kit (Millipore) according to the manufacturers instructions.

A.3.8 Ligation

Typical ligation reactions contained 50ng cut vector, 3-5 fold molar excess of insert and 40 units of T4 DNA ligase to a total volume of 10µl of 1 x T4 DNA ligase buffer.

Controls with cut vector with and without ligase were included. The reaction was left to incubate at room temperature for one hour and used to transform DH5α MAX Efficiency bacterial cells (Gibco). Uncut vector was transformed in parallel to monitor transformation efficiency.

A.4 Protein Expression

A.4.1 *E.coli* host strains

Unless stated otherwise electrocompetent BL21-(DE3 RecA$^-$) (designated FB810) cells were used for protein expression.

A.4.2 Transformation of chemically competent *E.coli* cells

Cells were thawed on ice and 1-10ng of the transforming DNA was added to the cells and mixed gently. The cells and DNA were incubated on ice for 30 minutes. The cells were heat shocked for 45 seconds at 41°C and returned to ice for 2 minutes. 250µl of SOC buffer was added and the cells incubated at 37°C for 1 hour. Selection was carried out by spreading the mixture on to dried L-AGAR plates containing the appropriate antibiotic and the colonies were counted after an overnight incubation at 37°C.

A.4.3 Transformation of electrocompetent *E.coli* cells

50µl of electrocompetent cells (FB810) were thawed on ice and were mixed with 1-5ng of transforming DNA. The mixture was transferred to a cooled EquiBio 2mm electroporation cuvette. The cuvette was then placed in an electroporator (BioRad Gene Pulser^{TM}) and pulsed at 2.5V, 25µF and 200Ω. 1ml of SOC media was added to the cuvette immediately and then plated out on selective AGAR media. The plates were incubated at 37°C overnight.

A.4.4 Bacterial Growth Measurement

E. coli growth was monitored by measuring the absorbance at 600nm in a Pharmacia Biotech Ultrospec2000 spectrometer compared to a media blank.

A.4.5 Expression of recombinant protein in *E.coli*

A single colony from a fresh transformation was inoculated into 100ml of LB media containing the correct antibiotic for the transforming DNA and grown overnight in a shaking incubator at 37°C and 220 rpm. The saturated overnight cultures were harvested by centrifugation in a Beckman GS-6R at 3000g for 15 minutes and the supernatant discarded to minimise transfer of β-lactamase to the large culture. The pellet was then re-suspended in fresh LB media and transferred to a 2l baffled flask containing 1l LB media with appropriate antibiotics. The flasks were incubated at 37°C and 175 rpm until an OD_{600} of 0.4 was reached. If the cells were to be induced at a lower temperature the flasks were transferred to either 18°C or 25°C and allowed to grow to an OD_{600} of 0.6. IPTG was added to a final concentration of 250μM and the cells allowed to grow for 3 hours at 37°C, 4 hours at 25°C or overnight at 18°C. The cells were harvested in a Beckman J-6B centrifuge at 4000g for 20 minutes. The cell pellets were re-suspended in 35ml of LB and harvested by centrifugation in a Beckman GS-6R at 3000g for 15 minutes, the supernatant discarded and the pellet stored at –80°C.

For expression trials alternative media was used to optimise protein expression. When TB media was used 100ml of potassium salts were added to 900ml of TB media. When NZY buffer was used 100ml of potassium salts and 4ml of casamino acids were added to 896 ml of NZY media.

A.4.6 Expression of Selenomethionine labelled recombinant protein in *E. coli*

Expression of Selenomethionine (Se-Met) labelled proteins was performed in methionine auxotroph cells, B834(DE3) cells (Novagen). For the autoclavable portion of the media for large scale production of Se-Met labelled protein 23.02g of amino acids (not including methionine) were dissolved to a total volume of 800ml in 22.25mM NaOH and the solution was autoclaved. The non-autoclavable portion consisted of 10g glucose, 0.5g $MgSO_4.7H_2O$, 10μl concentrated H_2SO_4 and 50g DL-SeMet to a total volume of 100ml in 1μM $FeCl_3$, 1μM thiamine and 0.05μM $CaCl_2$ and filter sterilised at 0.45μm. These two components were mixed and supplemented with 100ml 10x M9 salts and appropriate antibiotics to generate the complete medium.

The media for overnight cultures was the same as above but the Se-Met is replaced by 0.3mM final concentration of methionine.

A fresh colony of B834 competent cells transformed with the construct of interest was suspended in 5ml of the media for overnight cultures as described above and incubated overnight at 37 °C and 220 rpm. All 5ml of the overnight culture was used to inoculate 0.5l cultures and these were grown as for wild type protein.

A.5 Protein purification and analysis

A.5.1 SDS-PAGE

The SDS-PAGE method used was based on Laemmli (Laemmli 1970) and carried out basically as described by Sambrook *et al* (Sambrook *et al.* 1989). 12% resolving gel was used to visualise all proteins described in this thesis and consisted of 12% (w/v) acrylamide, 0.3% (w/v) bisacrylamide, 0.3M Tris-HCl pH 8.8, 0.1% (w/v) SDS, 0.003% TEMED and 0.0003% (w/v) APS. The 3.8% stacking gel consisted of 3.8% (w/v) acrylamide, 0.1% (w/v) bisacrylamide, 125mM Tris-HCl pH 6.8, 0.1% (w/v) SDS, 0.003% TEMED and 0.0003% (w/v) APS.

The TEMED and APS were added last as they induce rapid acrylamide polymerisation. Gels (plates 10 x 8 cm) were cast and run using apparatus purchased from Amersham Pharmacia Biotech. Samples to be analysed by SDS-PAGE were re-suspended in SDS-sample buffer and incubated at 95°C for 2 minutes. Bead-bound proteins were eluted in SDS-sample buffer by boiling for 5 minutes. Gels were run in 1 x SDS-PAGE running buffer with a limiting voltage of 100 V at the beginning of the run and 200 V during the second half of the run. Electrophoresis was continued until the bromophenol blue dye front had just run off the bottom of the gel.

Gels were stained at room temperature for at least 5 minutes with Coomassie brilliant blue (0.1% Coomassie blue R350, 30% (v/v) methanol, 10% (v/v) glacial acetic acid). Gels were destained (45 % (v/v) methanol and 9 % (v/v) glacial acetic acid) at room temperature.

A.5.2 Silver stain

Silver staining was carried out when visualisation by Coomassie brilliant blue was not possible using a silver stain kit from OWL Separation Systems according to the manufacturers instructions.

A.5.3 Purification of His-tagged proteins

Sonication Buffer: 300mM NaCl, 50mM Tris-HCl pH 8.0, 10mM Benzamadine, 5mM 2-ME, 1mM PMSF.

Wash Buffer: 300mM NaCl, 20mM Tris-HCl pH 8.0, 5mM 2-ME

Elution Buffer: 20mM Tris-HCl pH 8.0, 100mM Imidazole, 5mM 2-ME

Storage Buffer: 20mM Tris-HCl pH 8.0, 5mM 2-ME

Frozen 1 litre cell pellets were defrosted at room temperature and homogenised using a wide bore pipette in 22ml of sonication buffer. The cells were sonicated on ice using a MSE Soniprep 150 sonicator at 14.5μm for 1 min and then 1.5 minutes with cooling and end over end mixing in-between. Cell debris was pelleted using a Beckman Allegra 64R centrifuge at 29000g and 4°C for 30 minutes. The supernatant was mixed on a roller shaker at 4°C for 1.5 hours with 1 ml equilibrated Ni-NTA resin (Qiagen) per litre of culture. The resin was then spun down in a Beckman GS-6R at 600g and 4°C for 5 minutes and washed thoroughly prior to elution. The protein was then dialysed overnight into its storage buffer.

Depending on the vector used affinity cleavage protease was added and allowed to cleave at 4°C overnight. SDS-PAGE was used to monitor the efficiency of cleavage.

A.5.4 Purification of GST-tagged proteins

Sonication Buffer: 100mM NaCl, 50mM Tris-HCl pH 8.0, 10mM Benzamadine, 1mM DTT, 1mM PMSF

Wash Buffer 1: 100mM NaCl, 50mM Tris-HCl pH 8.0, 1mM DTT

Wash Buffer 2: 20mM Tris-HCl pH 8.0, 1mM DTT

Storage Buffer: 20mM Tris-HCl pH 8.0, 1mM DTT

The cell pellet was re-suspended and treated as described for His-tagged protein purification to yield a soluble cell lysate containing the GST-fusion protein of interest.

The cell lysate was added to 100µl glutathione Sepharose matrix (Amersham Biosciences) per 1 of initial culture. The lysate and resin were mixed together on a roller shaker at 4°C for 1.5 hours. The resin was spun down in a Beckman GS-6R at 600g and 4°C for 5 minutes and thoroughly washed in wash buffer 1 followed by wash buffer 2. The resin was re-suspended in 5ml of storage buffer.

Depending on the vector used affinity cleavage protease was added and allowed to cleave at 4°C overnight. Proteins remaining bound to the matrix following cleavage were eluted using SDS-sample buffer. SDS-PAGE was used to monitor the efficiency of cleavage by analysing both the proteins present in the supernatant and those remaining bound to the matrix following cleavage.

A.5.5 Protein Concentration

Protein concentrations were determined according to the method of Bradford (Bradford 1976) using Bio-Rad protein assay reagent according to the manufacturers instructions.

Proteins for crystal trials were concentrated to between 5 and 20mg/ml using an equilibrated Centricon concentrator (Amicon) in a Beckman GS-6R at 3700rpm and 4°C. The concentrated protein was either stored at 4°C for immediate use or in 50% glycerol at -80°C.

A.5.6 Gel filtration

Analytical and purification size exclusion chromatography was performed using the ÄKTA HPLC machine (Amersham Biosciences) using the Superdex 75 and Superdex 200 columns using the settings described by the manufacturer. Columns were equilibrated with 20mM Tris-HCl pH 8.0, 1mM DTT and 50mM NaCl and calibrated with proteins of a known mass. The calibration of the columns is not precise as this method depends on both the mass and the shape of the protein (the Stoke's radius) and so it is not possible to directly compare different proteins and obtain an accurate result despite being a good method for obtaining an initial estimate. Approximately 2mg of protein was loaded on to the column for analysis of the molecular weight and for purification the protein was concentrated to 500µl and the entire sample loaded.

A.6 Protein-protein interaction techniques

A.6.1 GST affinity pull-down

GST affinity pull-down experiments were performed by purifying various GST fusion proteins as described in 0 but terminating the protocol before proteolytic cleavage of the GST moiety and supplementing the storage buffer with 50mM NaCl to prevent non-specific binding. The GST fusion protein bound to the glutathione-Sepharose resin was used as bait for a variety of prey including purified recombinant proteins and cell lysates. These prey were incubated with the bound fusion protein for 1.5 hours at 4°C on a roller shaker. The supernatant was then removed and the resin thoroughly washed using the storage buffer with 50mM NaCl.

Depending on the vector used affinity cleavage protease was added and allowed to cleave at 4°C overnight.

The proteins were visualised by SDS-PAGE and bound proteins analysed by mass spectroscopy performed by Nick Totty (Protein Analysis Laboratory, Cancer Research UK) or by Western blot by Heike Laman (Wolfson Institute for Biomedical Research, UCL).

A.7 Biophysical techniques

A.7.1 Isothermal Titration Calorimetry

An ITC instrument consists of two identical cells composed of a highly efficient thermal conducting material surrounded by an adiabatic material. Heaters located on both cells and jackets are activated when necessary to maintain identical temperatures between the components. The macromolecule is placed in the sample cell and the reference cell is filled with water. During the injection of the titrant into the sample cell the time dependent input of power required to maintain equal temperatures in the sample and reference cell is recorded. The heat absorbed or evolved during a calorimetric titration is proportional to the fraction of bound ligand and so the accuracy of the initial concentrations of the macromolecule and titrant are vital to the experiment.

ITC was carried out with the help of Sarah Westcott (Molecular Enzymology Laboratory, Cancer Research UK) using the MicroCal ITC equipment according to

the manufacturers instructions and using methods described in Pierce *et al* (Pierce *et al.* 1999). All proteins analysed were dialysed into the same storage buffer consisting of 20mM Hepes pH 8.0, 50mM NaCl and 1mM 2-ME. Fbxo7-Skp1 complex purified by gel filtration was loaded into the cell and PI31-NTD injected by a syringe. Approximately 20μM of $Fbxo7^{F3}$-Skp1 was in the cell and 100μM of PI31 FP domain in the injectant. Analysis of the results was performed using UNICORN software provided by MicroCal using the one binding site model. Protein concentration was calculated using the NanoDrop spectrophotometer to accurately measure the OD_{280} of the protein, as precise protein concentration is critical for correct analysis of the ITC results.

All proteins analysed by ITC were purified by size exclusion chromatography and dialysed into the same buffer prior to measurement of their molar concentration. The experiments were carried out at 30°C with 28 injections each of 10μl measured with stirring and all settings were constant for both wild type and mutant experiments. Data corrected for heating of dilution were fitted using a nonlinear least square curve-fitting algorithm (MicroCal) with three floating variables: stoichiometry, binding constant and change in the enthalpy of interaction, which assumes one non-competitive binding site.

A.7.2 Analytical Ultracentrifugation

AUC experiments were performed with the help of Jane Sandall (Molecular Enzymology Laboratory, Cancer Research UK) on PI31-NTD and mutant proteins. All proteins were stored and analysed in a buffer containing 20mM Tris-HCl pH 8.0, 50mM NaCl and 5mM 2-ME. All AUC experiments were performed using a Beckman XL-I analytical ultracentrifuge using both absorbance and interference data according to the manufacturers instructions. Protein parameters used for data analysis and the calculation of initial experimental parameters were calculated using the program SEDNTERP (Laue *et al.* 1992). Data was analysed using the AUC machine software and the programs SEDFIT and SEDPHAT (Schuck 2000) using the model monomer-dimer equilibrium in the final analysis.

There are two different methods to obtain data from AUC experiments and each is needed to obtain a full characterisation of the protein. Firstly a sedimentation velocity experiment was carried out to obtain an initial characterisation of the purity of the

sample. This experiment consisted of the application of a centrifugal force and the analysis of the time-course of the sedimentation process. Secondly a sedimentation equilibrium experiment was performed to establish the thermodynamic interactions of the proteins using the application of a centrifugal force that permits the diffusion to balance the sedimentation such that a time-invariant equilibrium gradient can be observed.

The sedimentation process is governed by three factors: gravitational force, buoyancy and hydrodynamic friction and from the balance of these three forces the Svedberg equation is obtained (Equation A.1). This equation describes a relationship between the three directly measurable quantities for a single protein component: the sedimentation coefficient (obtained from the migration of the sedimentation boundary with time in a sedimentation velocity experiment), the diffusion coefficient (obtained from the spread of the sedimentation boundary with time in sedimentation velocity), and the molar mass (obtained from the exponential gradient in sedimentation equilibrium).

Equation A.1

$$\frac{s}{D} = \frac{M(1 - \bar{v}r)}{RT}$$

Where s is the sedimentation coefficient (measured in units of Svedberg, with 1 S = 10^{-13} sec), D is the diffusion constant, M is the protein molar mass, $\bar{v}$ is the effective protein partial-specific volume (which can be predicted from amino acid sequence (Laue *et al.* 1992)), r is the distance from the centre of rotation, R is the gas constant and T is the absolute temperature.

The major parameter obtained from the sedimentation velocity experiment is the sedimentation coefficient (s) described in Equation A.2.

Equation A.2

$$s = \frac{1}{\omega^2 r}\frac{dr}{dt} = \frac{m}{f}$$

Where s is the sedimentation coefficient, $\omega^2 r$ is the centrifugal field, r is the radius, t is time of sedimentation, m is buoyant molecular mass and f is the frictional coefficient of the protein.

The approach used to analyse the sedimentation velocity data was to describe it as a differential sedimentation coefficient distribution $c(s)$, which describes a superposition of sedimentation boundaries of many species with different s-values (Balbo and Schuck In Press). The distribution *c(s)* and residuals of the fit were used to assess if the $c(s)$ model was adequately describing the data. If the distribution

contained a single major peak, it was transformed to a molar mass distribution $c(M)$ that allowed estimation of the molar mass of the main species (Schuck 2000).

The theoretical description of reversible protein interactions is conceptually simple in sedimentation equilibrium conditions because chemical equilibrium and mechanical sedimentation equilibrium can be achieved simultaneously. This then requires only equilibrium thermodynamics to describe the model because the mass action law applies across the cell. In contrast, sedimentation velocity is governed by thermodynamics but also by the reaction kinetics and hydrodynamics making the theoretical description more complex (Balbo and Schuck In Press).

Sedimentation equilibrium experiments utilise the fact that at lower rotor speeds the transport of sample down the centrifuge cell is balanced by diffusion back up the cell due to the creation of a concentration gradient. Sedimentation equilibrium is established when no change in the concentration distribution of any component is detectable.

Appendix B X-ray crystallography methods and theory

B.1 Protein crystals

B.1.1 Crystallisation

The basic strategy for crystallisation is to bring a protein solute to a limited degree of supersaturation (Ducruix and Griegé 1992).

There are three discrete zones where the solution is supersaturated (Figure B.1):

1) Precipitation zone where the excess of protein molecules separate from solution to form amorphous aggregates.
2) Nucleation zone where the excess of protein molecules aggregate in a crystalline form.
3) Metastable zone where well ordered crystals of a large size can be grown providing nucleation has already taken place.

Two techniques for setting up crystal trials were used to find the optimum path through phase space (Figure B.1); vapour diffusion (either using sitting or hanging drops) and microbatch (either using paraffin oil (Chayen *et al.* 1990) or Al's oil (1:1 silicon and paraffin oil) (D'Arcy *et al.* 1996)).

In both sitting and hanging drop vapour diffusion experimental setups (Figure B.1a) typically a 1-5µl drop of sample was mixed with a 1-5µl drop of reservoir solution. The initial precipitant concentration in the drop was therefore lower than that in the reservoir. In order to reach equilibrium the reservoir draws water from the sample drop increasing the concentration of precipitant and protein solute in the drop. Both sitting drops and hanging drops sample the same area of phase space through the nucleation and metastable zones.

Microbatch experiments sample a different area of phase space to vapour diffusion (Ducruix and Griegé 1992) as equilibrium is reached rapidly. For all microbatch experiments 1µl of protein was mixed with 1µl of crystallization solution and placed under oil. When the drop is placed under paraffin oil (Figure B.1b) very little evaporation is possible and so the concentration of the protein is constant. When Al's oil is used (Figure B.1c) water can permeate the oil allowing evaporation and concentration of the sample over time.

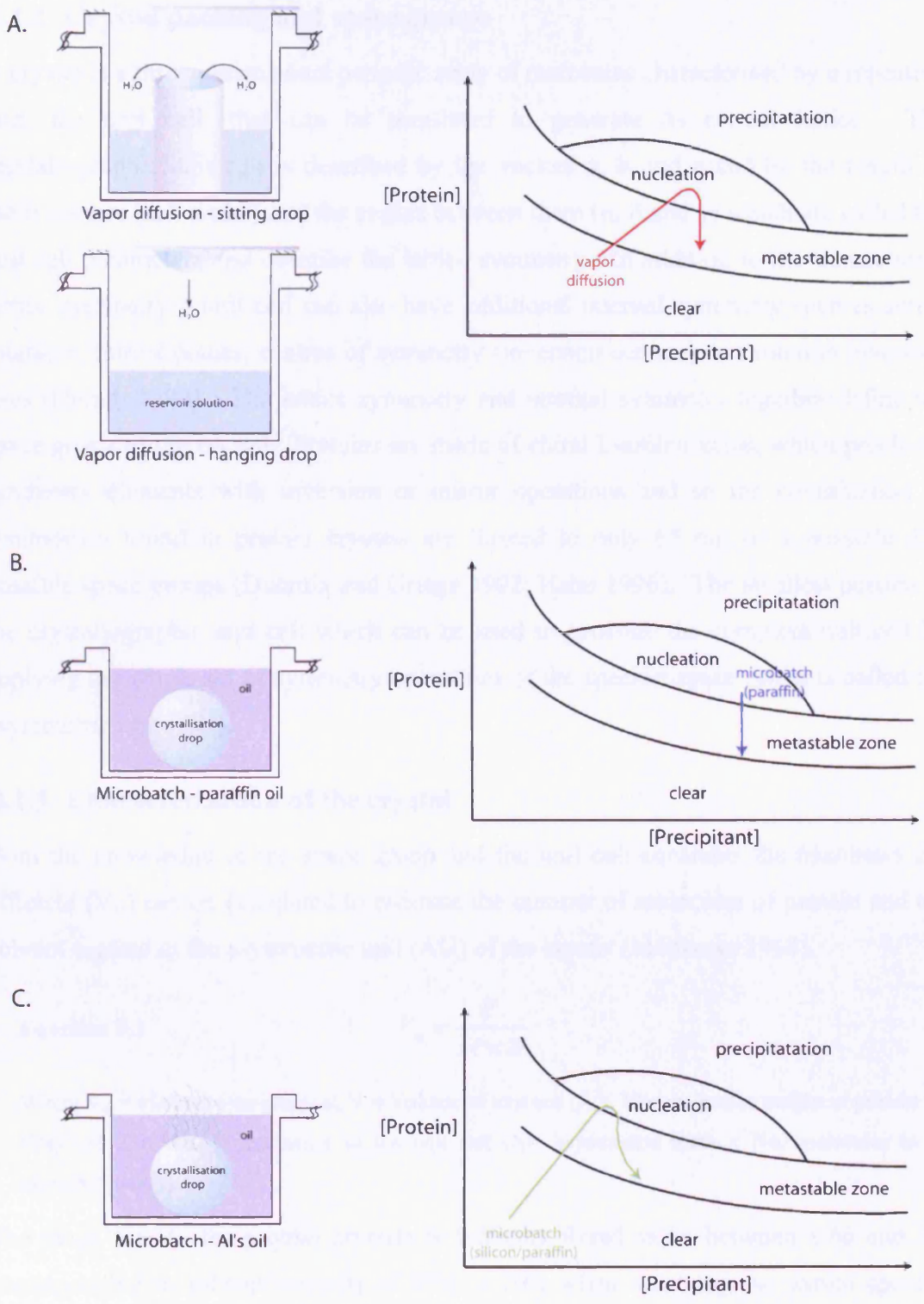

Figure B.1 - Phase diagram and crystallisation techniques

A. Vapour diffusion set-up for hanging and sitting drops. B. Microbatch using paraffin oil. C. Microbatch using Al's oil. Each point on the curves corresponds to a relative concentration at which the solution is in equilibrium with the precipitating agent. Arrows show the path through phase space for each type of experimental set-up. Set-up diagrams adapted from the Hampton website (http://www.hamptonresearch.com).

B.1.2 Crystal packing and space groups

A crystal is a three dimensional periodic array of molecules characterised by a repeating unit, the unit cell, that can be translated to generate its crystal lattice. The crystallographic unit cell is described by the vectors **a**, **b** and **c** and by the length of these vectors (a, b and c) and the angles between them (α, β and γ) which are called the unit cell parameters and describe the lattice symmetry. In addition to the translational lattice symmetry a unit cell can also have additional internal symmetry such as screw rotations, mirror planes, centres of symmetry (inversion centres) and rotation inversion axes (Drenth 1994). The lattice symmetry and internal symmetry together define the space group of the crystal. Proteins are made of chiral L-amino acids, which precludes symmetry elements with inversion or mirror operations and so the combination of symmetries found in protein crystals are limited to only 65 out of a possible 320 possible space groups (Ducruix and Griegé 1992; Hahn 1996). The smallest portion of the crystallographic unit cell which can be used to generate the complete unit cell by applying the whole set of symmetry operations of the specific space group is called the asymmetric unit (AU).

B.1.3 Characterisation of the crystal

With the knowledge of the space group and the unit cell constants the Matthews co-efficient (V_m) can be calculated to estimate the number of molecules of protein and the solvent content in the asymmetric unit (AU) of the crystal (Matthews 1968).

Equation B.1

$$V_m = \frac{V}{M \times Z}$$

Where V_m = Matthews co-efficient, V = Volume of unit cell (Å^3), M = molecular weight of protein (Da) and Z = No. of molecules in the unit cell (No. asymmetric units × No. molecules in asymmetric unit)

The range for V_m for protein crystals is typically found to be between 1.66 and 4.0 corresponding to solvent contents of 25% to 70% when assuming the partial specific volume of the protein is 0.74ml/g. Using this it is possible to calculate V_m assuming the number of molecules per AU = ¼, ½, 1, 2, 3 etc and from this narrow down the possible values. If V_m falls far outside the range then the number of molecules per AU assumed is likely to be incorrect because the packing of the molecules is either impossibly dense or too distant to form a 3D-lattice.

B.2 X-ray diffraction by crystals

B.2.1 How X-rays interact with a crystal lattice: Bragg's law and the Ewald construction

X-rays are scattered by the electrons within the atoms that make up proteins. A charged particle that interacts with an electromagnetic wave experiences a force, which causes it to oscillate. This causes it to act as a new source of electromagnetic radiation and so 'scatter' the incident radiation. This interaction can only be observed if the dimensions of the particle are of the same order of magnitude as the wavelength of the radiation; therefore X-ray radiation with a wavelength in the range of 1-1.5Å can be used to study the structure of matter at atomic resolution.

In order to observe constructive interference the X-rays must scatter in phase in a relationship described by Bragg's law (Equation B.2).

Equation B.2

$$2d\sin\theta = n\lambda$$

Where θ = angle of diffraction, λ = wavelength, n = integer value and d = inter-planar spacing between atoms

Bragg's law states that diffraction spots will only be observed at the angle of incidence at which the path length between the wave scattered from each plane is equal to an integer number of wavelengths. This relationship means that every atom contributes to every diffraction spot. The resolution of the X-ray diffraction pattern is the minimum inter-planar spacing (d_{min}) of a set of planes from which the diffraction can be observed.

Rearrangement of Bragg's law reveals that $\theta \propto (1/d)$ which means that it is easier to represent the diffraction data using a reciprocal lattice based on 1/d which varies as sinθ. This reciprocal lattice can be used to describe the unit cell edges as a/h, b/k and c/l where h, k and l are integers and are called the Miller indices. The notation (*hkl*) is used to describe the plane defined by the h, k and l Miller indices. The space described by the Miller indices is referred to as reciprocal space.

The Ewald construction is a theoretical method to help understand diffraction from a crystal. This construction uses the Bragg equation re-written in vector form (Equation B.3).

Equation B.3 $$\mathbf{S = s - s_0}$$

Where S is the scattering vector which also defines a reciprocal lattice point in the diffraction and $|s| = |s_0| = 1/\lambda$

Only when the vector triangle given by Equation B.3 and described in Figure B.2 is satisfied does diffraction occur. The Ewald sphere is only applicable if the direction of the beam is parallel to **s** for the orientation of the crystal. Providing these conditions are satisfied diffraction occurs whenever a point of the reciprocal lattice lies exactly on the Ewald sphere.

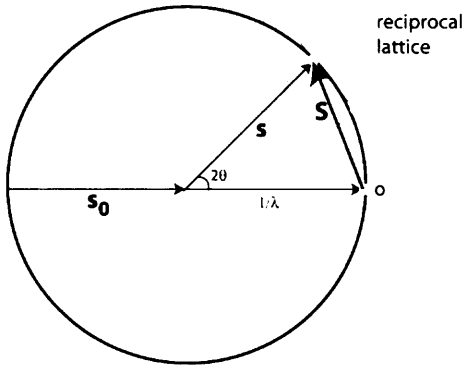

Figure B.2 – The Ewald sphere

The sphere has a radius $1/\lambda$, the centre of the reciprocal lattice is at O, s_0 indicates the direction of the incident beam and s indicates the direction of the scattered beam. This is a 2-dimensional representation of the sphere adapted from Drenth, J (Drenth 1994)

Each lattice point (reflection) can be characterised by the phase, amplitude and the position of the lattice point in reciprocal space. This relationship is described as the structure factor (F_{hkl}) and is described in Equation B.4. In order to interpret Equation B.4 for the position *hkl* the phase of each reflection must be known but this information cannot be measured directly and this is known as the phase problem.

Equation B.4 $$F_{hkl} = |F_{hkl}| \exp[i\alpha(hkl)]$$

Where $|F_{hkl}|$ is the structure factor amplitude and $\alpha(hkl)$ is its phase

Protein crystals have a unit cell much larger than the wavelength of the radiation used and so in order to measure data for all reflections the crystal is rotated to bring successive reflections into a diffracting position. The rotation method was employed for all data collection in this thesis and is where the crystal is rotated about a single axis

(φ) through a narrow, defined angle ($\Delta\varphi$) during each exposure to the X-rays. All reflections that satisfy the Bragg equation over this angle are recorded as diffraction spots on a single image.

B.2.2 X-ray sources and detectors

The X-rays used for protein crystal diffraction are typically generated by two different sources, either rotating anodes or synchrotrons.

Rotating anodes produce X-rays by focusing an accelerated beam of electrons in a vacuum onto a metal anode, typically copper. When the beam hits the metal most of the energy is liberated as heat, however some electrons will absorb a small amount of the energy and move to a higher energy state. When the excited electrons return from the higher orbitals to their normal energy state the energy is released in the form of X-ray radiation of a characteristic wavelength. The wavelength is dependent on the two orbitals between which the transition of the electron occurs. There are typically two wavelengths emitted from a copper anode (K_α=1.54 Å and K_β=1.39 Å). The K_β peak is filtered out and K_α radiation focussed, using mirrors, onto the protein crystal (Drenth 1994).

Synchrotron radiation sources produce X-rays by utilising the fact that charged particles travelling at relativistic speeds emit radiation in the X-ray range when they undergo a change in direction. In a synchrotron electrons at 2-6 GeV are forced by dipole magnets to adopt a curved path causing a constant change in direction and therefore a constant source of X-rays (Sweet 1998). Synchrotrons produce white radiation from which specific wavelengths can be chosen using a monochromator. Helliwell (Helliwell 1997) discusses experimental considerations.

The intensities of all the diffracted X-rays need to be measured in order to determine a crystal structure. A detecting device is required to measure the position and intensity of the diffracted beams. Accurate measurement of the diffraction pattern was performed using either a MAR storage phosphor-based image plate or a charge-coupled device (CCD).

B.2.3 Cryo-crystallography

Exposure of protein crystals to X-ray radiation causes damage due to the interaction between the molecules in the crystal and the beam. Energy is dissipated in two ways: it

produces heat through the thermal vibrations of the molecules and it provides the energy to break the bonds between atoms. The mechanisms for producing reactive radicals are direct (i.e. damage to the poly-peptide bonds) or indirect (reactive H* or OH* formed by the destruction of a water molecule). The thermal energy within the crystal can allow these free radicals to diffuse through the crystal causing further damage. At cryo temperatures (100 K) these reactive particles are immobilised in the crystal and cause less secondary damage (Garman 1999). Garman and Doublié (Garman and Doublié 2003) discuss experimental consideration and procedures.

The high water content of protein crystals means that cryo-protectant agents (e.g. glycerol, PEG 400 and N-paratone) have to be employed to avoid the destructive effects of crystalline ice formation during freezing.

B.2.4 Data collection

During data collection two factors influencing the completeness of data must be considered: 1) the symmetry of the reciprocal lattice and 2) the accuracy of the measured reflections which depend on factors such as the exposure time, crystal diffracting power and characteristics of the detector and X-ray source. The redundancy of the dataset is also important because when a reflection is measured several times its value can be determined with greater accuracy. The number of times each reflection is measured is referred to as the multiplicity of the dataset.

While theoretically the diffracted X-rays should form a sharp spot on the detector in practice various factors cause it to have a finite size. For example the incident radiation has some divergence and all parts of the crystal are not in the same orientation as described in the Ewald construction. X-rays from rotating anodes are parallel within a small angle (0.2-0.4°), which is improved for synchrotron radiation (≈0.001°), and crystals are composed of small mosaic blocks slightly mis-oriented with respect to each other (Dauter 1999). The wavelength of the radiation is also imprecise and so all these effects cause the diffraction from a particular reflection to be spread over a range of crystal rotation. This means that some reflections come into a diffracting position during one exposure and finish during the next; these are known as partial reflections.

All of the considerations for a complete dataset were taken into account for each specific experimental setup in this thesis by using the STRATEGY option within the

Mosflm program (Leslie 1992) to calculate the initial oscillation φ angle, oscillation range and the number of images needed to obtain a complete dataset.

B.2.5 Processing data in Mosflm

X-ray diffraction data consists of many images each containing the intensity and position of many reflections. The aim of data processing is to integrate this raw data into a three-dimensional representation of the reciprocal lattice points with their measured intensities. Several program suites are available to process X-ray diffraction data; however in this thesis only Mosflm (Leslie 1992) was used.

The first stage in data processing is obtaining an accurate estimate for the unit cell and space group of the crystal, this is done using a process called autoindexing based on a one-dimensional Fast Fourier Transform (FFT) method (Powell 1999). Using several images can further improve this estimate by refining the cell parameters, the crystal to detector distance, parameter defining the orientation of the crystal with respect to the incident beam and the mosaic spread of the crystal independently of all detector parameters and distortions by a process called post-refinement. After calculating these parameters the whole dataset is integrated to produce a file containing a list of reflections and their intensities after applying a correction for the effects of polarization and the Lorentz correction (which is a geometrical correction dependent on data collection technique).

B.2.6 Scaling and analysing the data

The program SCALA (Evans 1993, 1997) was used to scale both reflections and datasets together; it also removes the effects of beam fluctuation and mechanical errors. SCALA produces scaled intensities for the unique indexed reflections and also outputs statistically useful values that can be used to analyse the data quality and the integration. Data quality can be judged using the statistical indicator R_{meas}, which is derived from a comparison of the intensities. This gives an average ratio of the spread of intensities of the multiply measured symmetry equivalent reflections to the estimated value of the reflection intensity (Equation B.5) (Diederichs and Karplus 1997).

Equation B.5

$$R_{meas} = \frac{\sum_h \sqrt{\frac{n_h}{n_h - 1}} \sum_i^{n_h} \left| \hat{I}_h - I_{h,i} \right|}{\sum_h \sum_i^{n_h} I_{h,i}}$$

Where n_h is the multiplicity, $\hat{I}_h = \frac{1}{n_h} \sum_i^{n_h} I_{h,i}$ and $I_{h,i}$ is intensity of the i^{th} measurement of the h reflection.

The program TRUNCATE was used to convert the averaged intensities into a file containing structure factors. The structure factors are not on an absolute scale and so the Wilson plot (French and Wilson 1978) is used to estimate both an absolute scale and temperature factor for a set of observed structure factors. It is assumed that all atoms in the cell have the same isotropic thermal motion and therefore that Equation B.6 is valid; however this assumption can only be used for data with a resolution of 3Å or higher.

Equation B.6

$$\ln\left(\frac{\sum |F_{hkl}|^2}{\sum f_0^{\,2}} \right) = -\ln C - 2B\left(\frac{\sin\theta}{\lambda} \right)^2$$

Where $|F_{hkl}|^2$ = observed intensities, f_0 = atomic scattering factor of each atom in the unit cell, C=linear scale factor and B=temperature factor

B.3 Phase determination by SIR and SAD

The electron density in a crystal can be obtained by calculating the Fourier summation in Equation B.7.

Equation B.7

$$\rho(x,y,z) = \frac{1}{V} \sum_{hkl} |F_{hkl}| \exp[-2\pi i(hx + ky + lz) + i\alpha_{hkl}]$$

Where $|F_{hkl}|$ = structure factor amplitude of reflection (hkl) including the temperature factor, V=volume of the unit cell, ρ = electron density, α = phase angle and x,y,z are the co-ordinates of the atom in the unit cell.

The structure factor amplitude can be obtained directly from diffraction data but no information is available for the phase angles. Two methods to solve the phase problem were used for the PI31-NTD structure determination; Single Isomorphous Replacement (SIR) which requires the attachment of heavy atoms to the protein molecules in the crystals and Single Anomalous Diffraction (SAD) which depend on the presence of strong anomalously scattering atoms within the protein.

One useful tool in obtaining experimental phase information is the Patterson function which is a Fourier summation that does not require any phase information and relies on the square of the structure factor amplitude as can be seen from Equation B.8.

Equation B.8 $$P(uvw)=\frac{1}{V}\sum_{hkl}|F_{hkl}|^2\cos[2\pi(hu+kv+lw)]$$

Where P(uvw) is the Patterson function of u,v,w relative co-ordinates in Patterson space

In general the Patterson function will have a zero value because the vector will be zero as one or both ends will be at zero electron density. It will only have positive value if both ends are at atomic positions. For N atoms in a unit cell there will be N^2 peaks in the Patterson map which is obviously a prohibitively large number to interpret for proteins however it can be used to find the location of a limited number of heavy atoms (Drenth 1994).

B.3.1 Heavy atom derivatives to solve the phase problem

SIR experiments require both data from a native protein crystal and a crystal of at least one heavy atom derivative, which is isomorphous to the native crystal (having the same cell dimensions and orientation of molecules). This isomorphism means that the intensity differences between the native and the derivative are due to the heavy atoms and not to any other changes within the crystal.

The known data in an SIR case are the amplitudes of $\mathbf{F_P}$ and $\mathbf{F_{PH}}$ (structure factors of the native protein and the heavy atom replacement protein respectively). Their phase angles are not yet known; however for each reflection there is a difference in $|F|$ described in Equation B.9 and Figure B.3.

Equation B.9 $$\Delta|F|_{iso}=|F_{PH}|-|F_P|$$

This means that the Patterson map with the coefficients $(\Delta|F|_{iso})^2$ will give a map of the heavy atom arrangement in the unit cell.

The high solvent content of protein crystals means that solutions of heavy atoms can be soaked into the protein crystal to produce derivatives. This is a trial and error process with variables including solubility of heavy atom compounds in the crystallisation solution, soaking time and concentration of the heavy atom compound (Ducruix and Griegé 1992).

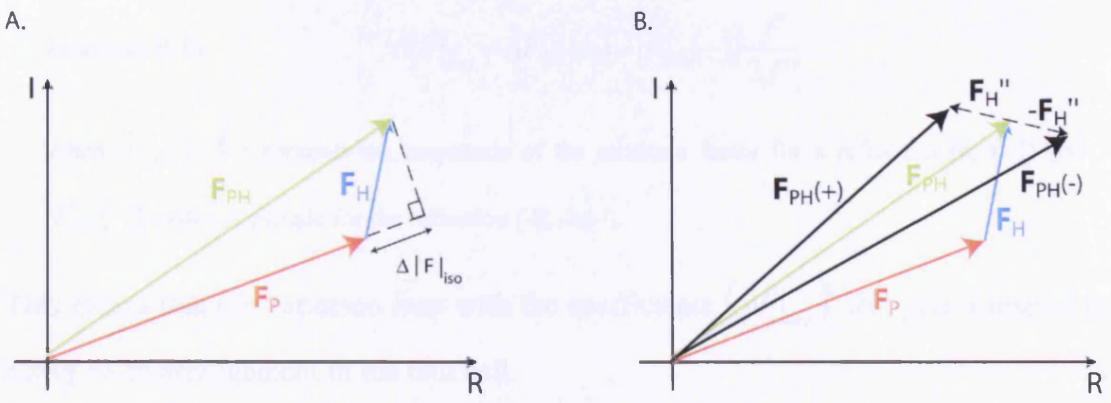

Figure B.3 – Argand diagram showing structure factors for SIR and SAD

A. The structure factor triangle for isomorphous replacement. B. Structure factor triangles for anomalous scattering, the structure factors $F_P(-)$, $F_{PH}(-)$, $F_H(-)$ and $F''_H(-)$ have been reflected with respect to the horizontal axis and combined with the structure factors for reflection (h, k, l). These diagrams have been reproduced from Drenth (Drenth 1994).

B.3.2 Anomalous diffraction to solve the phase problem

Diffraction is caused by electrons surrounding the atomic nucleus and is represented by the atomic form factor $f^0(\theta)$. If X-rays provide electrons with the energy required to leave an energy shell an anomalous signal is observed which can be expressed as a complex number and is dependent on the wavelength ($f' + if''$). f' is proportional to the atomic absorption of X-rays and f' follows the derivative of this. The full atomic form factor is shown in Equation B.10.

Equation B.10

$$f(\theta,\lambda) = f^0(\theta) + f'(\lambda) + if''(\lambda)$$

The atomic absorption increases with atomic number and as the energy of the X-rays approach the resonant energy of excitation of orbital electrons. As the energy moves below such a value the absorption diminishes rapidly (Hendrickson and Ogata 1997).

When the atomic form factor is real $F(h,k,l) \equiv F(-h,-k,-l)$ and $\varphi(h,k,l) \equiv -\varphi(-h,-k,-l)$; this is known as Friedel's law. The effect of anomalous scattering is that this law no longer applies, and so differences between intensities of Friedel pairs can be used as a means to obtain the phase angle for each structure factor using the difference described in Equation B.11 and shown in Figure B.4.

Equation B.11

$$\Delta|F|_{ano} = \{|F_{PH}(+)| - |F_{PH}(-)|\}\frac{f'}{2f''}$$

Where $|F_{PH}(+)|$ represents the amplitude of the structure factor for a reflection (h, k, l) and $|F_{PH}(-)|$ is the amplitude for the reflection (-h, -k, -l).

This means that the Patterson map with the coefficients $(\Delta|F|_{ano})^2$ will give a map of the heavy atom arrangement in the unit cell.

A multiple anomalous diffraction (MAD) dataset contains data from the same crystal at, usually, three different wavelengths, the peak, the inflection and the remote wavelength (Figure B.4) in order to maximise the information from obtained from the anomalous signal (Hendrickson and Ogata 1997). Single anomalous diffraction (SAD) datasets are generally collected at the peak wavelength where time, radiation damage or other limitations prevent a MAD dataset from being collected. SAD data can be used to obtain phases; however the data must be of good quality to ensure an accurate anomalous signal (Dauter *et al.* 2002) and additional methods such as density modification are required to overcome the phase ambiguity (Dodson 2003).

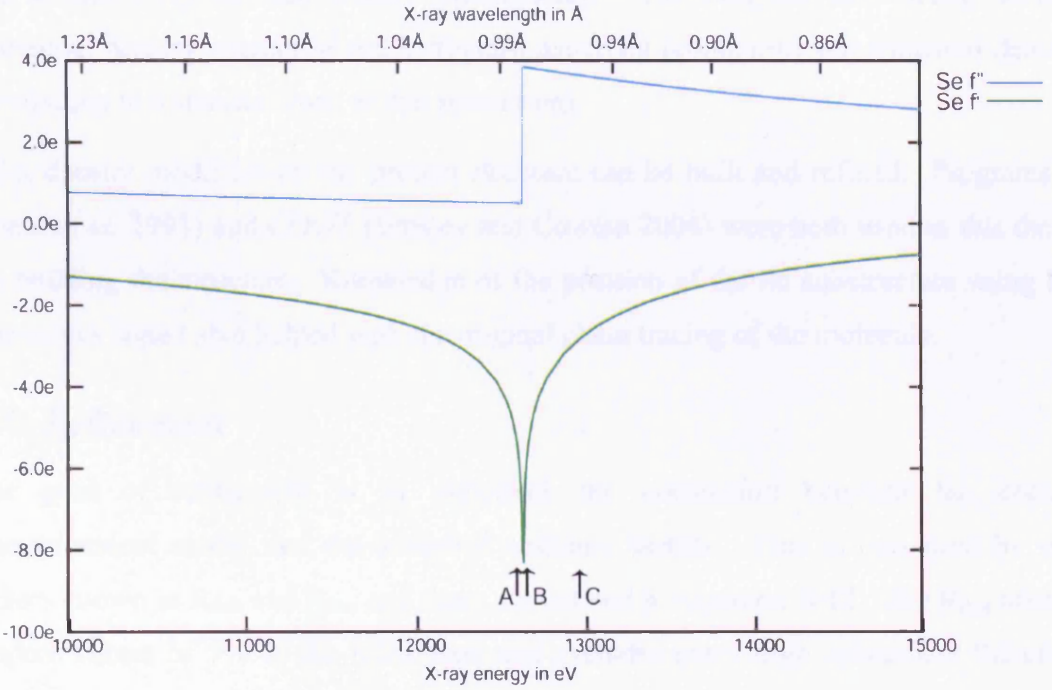

Figure B.4 – Theoretical *f'* and *f''* plot for Se K edge

Plotted using the server http://www-ssrl.slac.stanford.edu/absorb.html (Brennan and Cowan 1992). Arrows mark the position of the wavelengths for: A) inflection B) peak and C) High energy remote.

Identification of the heavy atom sites and estimation of phases from either isomorphous replacement or anomalous data were performed using SOLVE (Terwilliger and Berendzen 1999) or SHELXD (Schneider and Sheldrick 2002) using automated Patterson searching, cross-vector matching and quality control methods. All work using the program SHELX reported in this thesis was performed by Judith Murray-Rust (Structural Biology Laboratory, Cancer Research UK).

Having obtained experimental amplitudes and experimental phases based on a subset of atoms it is theoretically possible to produce an electron density map. However in practice early electron density maps from SIR and anomalous scattering are very noisy, and so density modification is applied to recognise and reduce errors in the phase estimate. When the map has been corrected phases are re-calculated and re-combined with the experimental phases. This can yield a better map and can also be used to solve the phase ambiguity in a SAD or SIR experiment. The programs RESOLVE (Terwilliger 2000) and DMMULTI (Cowtan 1994) were used in this thesis to perform density modification and solvent flattening. RESOLVE uses a maximum likelihood approach to density modification while DMMULTI allows data to be combined from two different crystals for density modification. The program uses Wang solvent flattening, density averaging and histogram matching (comparing experimental density distribution to a standard map at that resolution).

After density modification the protein structure can be built and refined. Programs O (Jones *et al.* 1991) and COOT (Emsley and Cowtan 2004) were both used in this thesis for building the structure. Knowledge of the position of the Se substructure using the anomalous signal also helped with the original chain tracing of the molecule.

B.4 Refinement

The goal of refinement is to maximise the correlation between the correct stereochemical model and the observed structure factors. This is measured by two factors known as R_{fact} and R_{free} and these are defined in Equation B.12. The R_{free} takes a random subset of 5% of the reflections and excludes them from refinement therefore offering a method of cross-validation (Brunger 1992).

Equation B.12

$$R_{fact} = \frac{\sum_{hkl \subset T} \left\| F_{obs} \right| - k \left| F_{calc} \right\|}{\sum_{hkl \subset T} \left| F_{obs} \right|} \qquad R_{free} = \frac{\sum_{hkl \subset T} \left\| F_{obs} \right| - k \left| F_{calc} \right\|}{\sum_{hkl \subset T} \left| F_{obs} \right|}$$

Where T is the subset of reflections used for cross validation

Refinement programs use a target function in order to obtain the best possible model. The program used for refinement in this thesis was REFMAC 5.0 (Murshudov *et al.* 1997) which employs a maximum likelihood function using Bayes' theorem which says that the best model is that most consistent with the observations. Restrained refinement was performed in this thesis, which means that the *a priori* information about macromolecular structures (bond length, bond angles, planarity etc) was used as additional information within the refinement.

At the low resolution (2.64Å) of PI31-NTD there were not enough observations per parameter to perform anisotropic refinement using anisotropic displacement parameters (ADPs) however these parameters can be modelled as those of a set of rigid bodies using TLS parameters (Winn *et al.* 2001; Winn *et al.* 2003). TLS parameters model the translation, rotation and screw-rotation of a rigid body and output a matrix that takes into account anisotropic behaviour. The output of TLS refinement in REFMAC was analysed using the program TLSANL (Howlin *et al.* 1993).

References

Aguilar, R. C. and Wendland, B. (2003). "Ubiquitin: not just for proteasomes anymore." Current Opinion in Cell Biology **15**: 184-190.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J. D. (1989). Molecular Biology of The Cell. New York & London, Garland Publishing, Inc.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Research **25**(17): 3389-3402.

Appelbaum, E. R. and Shatzman, A. R. (1999). Prokaryotic *in vivo* expression systems. Protein Expression. S. J. Higgins and B. D. Hames, Oxford University Press: 169-199.

Baba, D., Maita, N., Jee, J. G., Uchimura, Y., Saitoh, H., Sugasawa, K., Hanaoka, F., Tochio, H., Hiroaki, H. and Shirakawa, M. (2005). "Crystal structure of thymine DNA glycosylase conjugated to SUMO-1." Nature **435**: 979.

Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J. W. and Elledge, S. J. (1996). "SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box." Cell **86**(2): 263-74.

Balbo, A. and Schuck, P. (In Press). Analytical Ultracentrifugation in the Study of Protein Selfassociation and Heterogeneous Protein-Protein Interactions. Protein-Protein Interactions: A Molecular Cloning Manual. E. Golemis and P. D. Adams.

Bassal, S., Nomura, N., Venter, D., Brand, K., McKay, M. J. and van der Spek, P. J. (2001). "Characterization of a Novel Human Cell-Cycle-Regulated Homologue of *Drosophila dlg1*." Genomics **77**(1-2): 5-7.

Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L. L., Studholme, D. J., Yeats, C.

and Eddy, S. R. (2004). "The Pfam protein families database." Nucl. Acids Res. **32**(90001): D138-141.

Bjellqvist, B., Hughes, G. J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J. C., Frutiger, S. and Hochstrasser, D. (1993). "The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences." Electrophoresis **14**(10): 1023-31.

Bochtler, M., Ditzel, L., Groll, M., Hartmann, C. and Huber, R. (1999). "The Proteasome." Annu Rev Biophys Biomol Struct **28**: 295-317.

Bornstein, G., Bloom, J., Sitry-Shevah, D., Nakayama, K., Pagano, M. and Hershko, A. (2003). "Role of the SCF^{Skp2} Ubiquitin Ligase in the Degradation of $p21^{Cip1}$ in S Phase." Journal Biological Chemistry **278**(28): 25752-25757.

Bradford, M. M. (1976). "A Rapid and Sensitive Method for the Quantitation of Microgram Quantities of Protein Utlilizing the Principle of Protein-Dye Binding." Analytical Biochemistry **72**: 248-254.

Brannigan, J. A., Dodson, G., Duggleby, H. J., Moody, P. C. E., Smith, J. L., Tomchick, D. R. and Murzin, A. G. (1995). "A protein catalytic framework with an N-terminal nucleophile is capable of self-activation." Nature **378**: 416-419.

Brennan, S. and Cowan, P. L. (1992). "A suite of programs for calculating x-ray absorption, reflection and diffraction performance for a variety of materials at arbitrary wavelengths." Rev. Sci. Instrum. **63**: 850.

Brunger, A. T. (1992). "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures." Nature **355**(6359): 472.

Carrano, A. C., Eytan, E., Hershko, A. and Pagano, M. (1999). "Skp2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27." Nature Cell Biology **1**: 193-199.

Carugo, K. D., Helliwell, J. R., Stuhrmann, H. and Weiss, M. S. (2005). "Softer and soft X-rays in macromolecular crystallography." Journal of Synchrotron Radiation **12**: 410-419.

Castro, A., Bernis, C., Vigeron, S., Labbé, J.-C. and Lorca, T. (2005). "The anaphase-promoting complex: a key factor in the regulation of cell cycle." Oncogene **24**: 314-325.

Cenciarelli, C., Chiaur, D. S., Guardavaccaro, D., Parks, W., Vidal, M. and Pagano, M. (1999). "Identification of a family of human F-box proteins." Curr Biol **9**(20): 1177-9.

Chayen, N. E., Stewart, P. D. S., Maeder, D. L. and Blow, D. M. (1990). "An Automated System for Micro-Batch Protein Crystallization and Screening." J. Appl. Cryst. **23**: 297-302.

Chiu, A. W., Huang, Y.-L., Huan, S. K., Wang, Y.-C., Ju, J.-P., Chen, M.-F. and Chou, C.-K. (2002). "Potential molecular marker for detecting transitional cell carcinoma." Urology **60**(1): 181-185.

Cho, S., Lee, M. G., Yang, J. K., Lee, J. Y., Song, H. K. and Suh, S. W. (2000). "Crystal structure of *Esherichia coli* CyaY protein reveals a previously unidentified fold for the evolutionarily conserved frataxin family." Proc Natl Acad Sci U S A **97**(16): 8932-8937.

Chu-Ping, M., Slaughter, C. A. and DeMartino, G. N. (1992). "Purification and characterization of a protein inhibitor of the 20S proteasome (macropain)." Biochemica et Biophysica Acta **1119**: 303-311.

Ciechanover, A. and Ben-Saadon, R. (2004). "N-terminal ubiquitination: more protein substrates join in." Trends in Cell Biology **14**(3): 103-106.

Ciechanover, A., Heller, H., Elias, S., Haas, A. L. and Hershko, A. (1980). "ATP-dependent conjugation of reticulocyte proteins with the polypeptide

required for protein degradation." Proc Natl Acad Sci U S A **77**(3): 1365-1368.

Cohen, S. N., Chang, A. C., Boyer, H. W. and Helling, R. B. (1973). "Construction of biologically functional bacterial plasmids in vitro." Proc Natl Acad Sci U S A **70**(11): 3240-4.

Collaborative Computational Project (1994). "The CCP4 Suite: Programs for Protein Crystallography." Acta Crystallogr D Biol Crystallogr **D50**: 760-763.

Cowtan, K. (1994). "'dm': An automated procedure for phase improvement by density modification." Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography **31**: 34-38.

Cudney, R., Patel, S., Weisgraber, K., Newhouse, Y. and McPherson, A. (1994). "Screening and optimization strategies for macromolecular crystal growth." Acta Crystallogr D Biol Crystallogr **50**(Pt 4): 414-23.

D'Arcy, A., Elmore, C., Stihle, M. and Johnston, J. E. (1996). "A novel approach to crystallising proteins under oil." Journal of Crystal Growth **168**: 175-180.

Dauter, Z. (1999). "Data-collection strategies." Acta Crystallogr D Biol Crystallogr **55 (Pt 10)**: 1703-17.

Dauter, Z. and Dauter, M. (2001). "Entering a New Phase: Using Solvent Halide Ions in Protein Structure Determination." Structure **9**: R21-R26.

Dauter, Z., Dauter, M. and Dodson, E. (2002). "Jolly SAD." Acta Crystallogr D Biol Crystallogr **D58**: 494-506.

Dauter, Z., Dauter, M. and Rajashankar, K. R. (2000). "Novel approach to phasing proteins: derivatization by short cryo-soaking with halides." Acta Crystallogr D Crystallogr **D56**: 232-237.

DeLano, W. L. (2004). The PyMOL Molecular Graphics System. San Carlos, CA.

Deshaies, R. J. (1999). "SCF and Cullin/RING H2-based ubiquitin ligases." Annu Rev Cell Dev Biol **15**: 435-467.

Deveraux, Q., Ustrell, V., Pickart, C. M. and Rechsteiner, M. (1994). "A 26 S Protease Subunit That Binds Ubiquitin Conjugates." Journal Biological Chemistry **269**(10): 7059-7061.

Diederichs, K. and Karplus, P. A. (1997). "Improved R-factors for diffraction data analysis in macromolecular crystallography." Nat Struct Biol **4**(4): 269-75.

Dodson, E. (2003). "Is it jolly SAD?" Acta Crystallogr D Biol Crystallogr **59**(Pt 11): 1958-65.

Drenth, J. (1994). Principles of Protein X-ray Crystallography, Springer-Verlag.

Driscoll, J., Brown, M. G., Finley, D. and Monaco, J. J. (1993). "MHC-linked *LMP* gene specifically alter peptidase activity of the proteasome." Nature **365**: 262-264.

Ducruix, A. and Griegé, R. (1992). Crystallization of Nucleic Acids and Proteins, IRL Press.

Elsasser, S., Chandler-Militello, D., Müller, B., Hanna, J. and Finley, D. (2004). "Rad23 and Rpn10 Serve as Alternative Ubiquitin Receptors for the Proteasome." Journal Biological Chemistry **279**(26): 26817-26822.

Emsley, P. and Cowtan, K. (2004). "Coot: Model-Building Tools for Molecular Graphics." Acta Crystallogr D Biol Crystallogr **D60**: 2126-2132.

Evans, P. R. (1993). Data Reduction. Proceedings of CCP4 Study Weekend on Data Reduction and Processing.

Evans, P. R. (1997). Scaling of MAD data. Proceedings of CCP4 Study Weekend on Advances in Phasing.

Evans, T., Rosenthal, E. T., Youngblom, J., Distel, D. and Hunt, T. (1983). "Cyclin: A Protein Specified by Maternal mRNA in Sea Urchin Eggs That Is Destroyed at Each Cleavage Division." Cell **33**: 389-396.

Flick, K., Ouni, I., Wohlschlegel, J. A., Capati, C., McDonald, W. H., Yates, J. R. r. and Kaiser, P. (2004). "Proteolysis-independent regulation of the transcription factor Met4 by a single Lys 48-linked ubiquitin chain." Nature Cell Biology **6**(7): 634-640.

Förster, A., Masters, E. I., Whitby, F. G., Robinson, H. and Hill, C. P. (2005). "the 1.9 Å Structure of a Proteasome-11S Activator Complex and Implications for Proteasome-PAN/PA700 Interactions." Mol Cell **18**: 589-599.

Förster, A., Whitby, F. G. and Hill, C. P. (2003). "The pore of activated 20S proteasomes has an ordered 7-fold symmetric conformation." Embo J **22**(17): 4356-4364.

Freed, E., Lacey, K. R., Huie, P., Lyapina, S. A., Deshaies, R. J., Stearns, T. and Jackson, P. K. (1999). "Components of an SCF ubiquitin ligase localize to the centrosome and regulate the centrosome duplication cycle." Genes Dev **13**(17): 2242-57.

French, S. and Wilson, K. (1978). "On the Treatment of Negative Intensity Observations." Acta Crystallogr A Crystallogr **A34**: 517-525.

Früh, K., Gossen, M., Wang, K., Bujard, H., Peterson, P. A. and Yang, Y. (1994). "Displacement of housekeeping proteasome subunits by MHC-encoded LMPs: a newly discovered mechanism for modulating the multicatalytic proteinase complex." Embo J **13**(14): 3236-3244.

Fujiwara, K., Tenno, T., Sugasawa, K., Jee, J. G., Ohki, I., Kojima, C., Tochio, H., Hiroaki, H., Hanaoka, F. and Shirakawa, M. (2004). "Structure of the ubiquitin-interacting motif of S5a bound to the ubiquitin-like domain of HR23B." J Biol Chem **279**(6): 4760-7.

Gaczynska, M., Rock, K. L. and Goldberg, A. L. (1993). "γ-Interferon and expression of MHC genes regulate peptide hydrolysis by proteasomes." Nature **365**: 264-267.

Garman, E. (1999). "Cool data: quantity AND quality." Acta Crystallogr D Biol Crystallogr **55 (Pt 10)**: 1641-53.

Garman, E. and Doublié, S. (2003). "Cryocooling of Macromolecular Crystals: Optimisation Methods." Methods in Enzymology **368**: 188-216.

Gassner, N. C. and Matthews, B. W. (1999). "Use of differentially substituted selenomethionine proteins in X-ray structure determination." Acta Crystallogr D Biol Crystallogr **D55**: 1967-1970.

Glickman, M. H. and Ciechanover, A. (2002). "The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction." Physiol Rev **82**: 373□428.

Glickman, M. H., Rubin, D. M., Coux, O., Wefes, I., Pfeifer, G., Cjeka, Z., Baumeister, W., Fried, V. A. and Finley, D. (1998). "A Subcomplex of the Proteasome Regulatory Particle Required for Ubiquitin-Conjugate Degradation and Related to the COP9-Signalosome and eIF3." Cell **94**: 615-623.

Glotzer, M., Murray, A. W. and Kirschner, M. W. (1991). "Cyclin is degraded by the ubiquitin pathway." Nature **349**: 132-138.

Goldberg, A. L. and Rock, K. L. (2002). "Not just research tools - proteasome inhibitors offer therapeutic promise." Nature Medicine **8**(4): 338-340.

Gouet, P., Courcelle, E., Stuart, D. I. and Metoz, F. (1999). "ESPript: multiple sequence alignements in PostScript." Bioinformatics **15**: 305-308.

Groll, M. and Clausen, T. (2003). "Molecular shredders: how proteasomes fulfill their role." Current Opinion in Structural Biology **13**: 665-675.

Groll, M., Ditzel, L., Löwe, J., Stock, D., Bochtler, M., Bartunik, H. D. and Huber, R. (1997). "Structure of 20S proteasome from yeast at 2.4Å resolution." Nature **386**: 463-471.

Grosse-Kunstleve, R. W. and Adams, P. D. (2003). "On symmetries of substructures." Acta Crystallogr D Biol Crystallogr **D59**: 1974-1977.

Guan, K. and Dixon, J. E. (1991). "Eukaryotic Proteins Expressed in Escherichia coli: An Improved Thrombin Cleavage and Purification Procedure of Fusion Proteins with Glutathione S-Transferase." Analytical Biochemistry **192**: 262-267.

Guardavaccaro, D. and Pagano, M. (2004). "Oncogenic aberrations of cullin-dependent ubiquitin ligases." Oncogene **23**: 2037-2049.

Haas, A. L. and Rose, I. A. (1982). "The Mechanism of Ubiquitin Activating Enzyme." Journal Biological Chemistry **257**(17): 10329-10337.

Haas, A. L., Warms, J. V. and Rose, I. A. (1983). "Ubiquitin adenylate: structure and role in ubiquitin activation." Biochemistry **22**(19): 4388-94.

Hahn, T. (1996). Space-Group Symmetry. Dordrecht/Boston/London, Kluwer Academic Publishers.

Hatakeyama, S. and Nakayama, K.-i. I. (2003). "U-box proteins as a new family of ubiquitin ligases." Biochem. Biophys. Res. Comm. **302**: 635-645.

Helliwell, J. R. (1997). "Overview of Synchrotron Radiation and Macromolecular Crystallography." Methods in Enzymology **276**: 203-217.

Hendrickson, W. A. and Ogata, C. M. (1997). "Phase Determination for Multiwavelength Anomalous Diffraction Measurements." Methods in Enzymology **276**: 494-523.

Hershko, A. and Ciechanover, A. (1998). "The ubiquitin system." Annu Rev Biochem **67**: 425-79.

Hershko, A., Ciechanover, A., Heller, H., Haas, A. L. and Rose, I. A. (1980). "Proposed role of ATP in protein breakdown: Conjugation of proteins with multiple chains of the polypeptide of ATP-dependent proteolysis." Proc Natl Acad Sci U S A **77**(4): 1783-1786.

Hershko, A., Heller, H., Elias, S. and Ciechanover, A. (1983). "Components of Ubiquitin-Protein Ligase System." J Biol Chem **258**(13): 8206-8214.

Hochstrasser, M. (2004). "Ubiquitin signalling: what's in a chain?" Nature **6**(7): 571-572.

Hod, D., Cenciarelli, C., Jensen, J. P., Nguyen, H. B. and Weissman, A. M. (1994). "Activation-dependent Ubiquitination of a T Cell Antigen Receptor Subunit on Multiple Intracellular Lysines." Journal Biological Chemistry **269**(19): 14244-14247.

Holm, L. and Sander, C. (1994). "Searching Protein Structure Databases Has Come Of Age." Proteins **19**: 165-173.

Hoppe, T. (2005). "Multiubiquitylation by E4 enzymes: 'one size' doesn't fit all." TRENDS in Biochemical Sciences **30**(4): 183-187.

Howlin, B., Butler, S. A., Moss, D. S., Harris, G. W. and Driessen, H. P. C. (1993). "TLSANL: TLS parameter-analysis program for segmented anisotropic refinement of macromolecular structures." Journal Applied Crystallography **26**: 622-624.

Hsu, J. M., Lee, Y. C., Yu, C. T. and Huang, C. Y. (2004). "Fbx7 functions in the SCF complex regulating Cdk1-cyclin B-phosphorylated hepatoma up-regulated protein (HURP) proteolysis by a proline-rich region." J Biol Chem **279**(31): 32592-602.

Huibregtse, J. M., Scheffner, M., Beaudenon, S. and Howley, P. M. (1995). "A family of proteins structurally and functionally related to the E6-AP ubiquitin-protein ligase." Proc Natl Acad Sci U S A **92**: 2563-2567.

Ilyin, G. P., Rialland, M., Pigeon, C. and Guguen-Guillouzo, C. (2000). "cDNA cloning and expression analysis of new members of the mammalian F-box protein family." Genomics **67**(1): 40-7.

Jackson, P. K., Elridge, A. G., Freed, E., Furstenthal, L., Hsu, J. Y., Kaiser, B. K. and Reimann, J. D. R. (2000). "The lore of the RINGS: substrate recognition and catalysis by ubiquitin ligases." Trends in Cell Biology **10**: 429-439.

Jancarik, J. and Kim, S. H. J. (1991). "Sparse matrix sampling: a screening method for crystallisation of proteins." Journal Applied Crystallography **24**: 409-411.

Jin, P., Hardy, S. and Morgan, D. O. (1998). "Nuclear Localization of Cyclin B1 Controls Mitotic Entry After DNA Damage." J. Cell Biol. **141**(4): 875-885.

Jones, S. and Thornton, J. M. (1995). "Protein-protein interaction: a review of protein dimer structures." Prog Biophys Mol Biol **63**(1): 31-65.

Jones, S. and Thornton, J. M. (1996). "Principles of protein-protein interactions derived from structural studies." Proc Natl Acad Sci U S A **93**: 13-20.

Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991). "Improved methods for building protein models in electron density maps and the location of errors in these models." Acta Crystallogr A Crystallogr **A47**: 110-119.

Kaiser, P., Flick, K., Wittenberg, C. and Reed, S. I. (2000). "Regulation of Transcription by Ubiquitination without Proteolysis: Cdc34/SCF^{Met30}-Mediated Inactivation of the Transcription Factor Met4." Cell **102**: 303-314.

Kamura, T., Hara, T., Kotoshiba, S., Yada, M., Ishida, N., Imaki, H., Hatakeyama, S., Nakayama, K. and Nakayama, K. I. (2003). "Degradation of p57^{Kip2} mediated by SCF^{Skp2}-dependent ubiquitylation." Proc Natl Acad Sci U S A **100**(18): 10231-10236.

Kamura, T., Koepp, D. M., Conrad, M. N., Skowyra, D., Moreland, R. J., Iliopoulos, O., Lane, W. S., Kaelin Jr, W. G., Elledge, S. J., Conaway, R. C., Harper, J. W. and

Conaway, J. W. (1999). "Rbx1, a Component of the VHL Tumour Suppressor Complex and SCF Ubiquitin Ligase." Science **284**: 657-661.

Kelley A.L., e. a. (2000). "Enhanced genome annotation using structural profiles in the program 3D-PSSM." Journal Molecular Biology **299**(2): 499-520.

Kesmir, C., van Noort, V., de Boer, R. J. and Hogeweg, P. (2003). "Bioinformatic analysis of functional differences between the immunoproteasome and the constitutive proteasome." Immunogenetics **55**: 437-449.

Kile, B. T., Schulman, B. A., Alexander, W. S., Nicola, N. A., Martin, H. M. E. and Hilton, D. J. (2002). "The SOCS box: a tale of destruction and degradation." TRENDS in Biochemical Sciences **27**(5): 235-241.

Kim, S. Y., Herbst, A., Tworkowski, K. A., Salghetti, S. E. and Tansey, W. P. (2003). "Skp2 Regulates Myc Protein Stability and Activity." Mol Cell **11**: 1177-1188.

Kloetzel, P.-M. (2004). "Generation of major histocompatibilty complex class I antigens: functional interplay between proteasomes and TppII." Nature Immunology **5**(7): 661-669.

Kofler, M., Motzny, K., Beyermann, M. and Freund, C. (In Press). "Novel Interaction Partners Of The CD2BP2-GYF Domain." Journal Biological Chemistry.

Kopp, F. and Kuehn, L. (2003). "Orientation of the 19 S Regulator Relative to the 20 S Core Proteasome: An Immunoelectron Microscopic Study." Journal Molecular Biology **329**: 9-14.

Koth, C. M., Orlicky, S. M., Larson, S. M. and Edwards, A. M. (2003). "Use of limited proteolysis to identify protein domains suitable for structural analysis." Methods Enzymol **368**: 77-84.

Krissinel, E. and Henrick, K. (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." Acta Crystallographica Section D **60**(12 Part 1): 2256-2268.

Krissinel, E. and Henrick, K. (2005). Protein structure comparison service SSM at European Bioinformatics Institute.

Krogh, A., Brown, M., Mian, I. S., Sjölander, K. and Haussler, D. (1994). "Hidden Markov Models in Computational Biology: Applications To Protein Modelling." Journal of Molecular Biology **235**: 1501-1531.

Krüger, E., Kloetzel, P.-M. and Enenkel, C. (2001). "20S proteasome biogenesis." Biochemie **83**: 289-293.

Laemmli, U. K. (1970). "Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4." Nature **227**: 680 - 685.

Laman, H., Funes, J. M., Ye, H., Henderson, S., Galinanes-Garcia, L., Hara, E., Knowles, P., McDonald, N. and Boshoff, C. (2005). "Transforming activity of Fbxo7 is mediated specifically through regulation of cyclin D/cdk6." Embo J **24**(1): 3104-3116.

Laskowski, R. A. (1995). "SURFNET: A program for visualising molecular surfaces, cavities and intermolecular interactions." J Mol Graph **13**: 323-330.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." Journal Applied Crystallography **26**: 283-291.

Laskowski, R. A., Watson, J. D. and Thornton, J. M. (2005). "ProFunc: a server for predicting protein function from 3D structure." Nucleic Acids Research **33**: W89-W93.

Laue, T. M., Shah, B. D., Ridgeway, T. M. and Pelletier, S. L. (1992). Analytical Ultracentrifugation in Biochemistry and Polymer Science. S. Harding and A. Rowe, Royal Society of Chemistry: 90-125.

Lee, M.-H. and Yang, H.-Y. (2003). "Regulators of G1 cyclin-dependent kinases and cancers." Cancer and Metastasis Reviews **22**: 435-449.

Leslie, A. G. W. (1992). "Recent changes to the MOSFLM package for processing film and image plate data." Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography **26**.

Lovering, R., Hanson, I. M., Borden, K. L. B., Martin, S., O'Reilly, N. J., Evan, G. I., Rahman, D., Pappin, D. J. C., Trowsdale, J. and Freemont, P. S. (1993). "Identification and preliminary characterisation of a protein motif related to the zinc finger." Proc Natl Acad Sci U S A **90**: 2112-2116.

Löwe, J., Stock, D., Jap, B., Zwickl, P., Baumeister, W. and Huber, R. (1995). "Crystal Structure of the 20S Proteasome from the Archaeon *T. acidophilum* at 3.4 Å Resolution." Science **268**: 533-539.

Matthews, B. W. (1968). "Solvent content of protein crystals." J Mol Biol **33**(2): 491-7.

McCutchen-Maloney, S. L., Matsuda, K., Shimbara, N., Binns, D. D., Tanaka, K., Slaughter, C. A. and DeMartino, G. N. (2000). "cDNA Cloning, Expression, and Functional Characterisation of PI31, a Proline-rich Inhibitor of the Proteasome." Journal Biological Chemistry **24**: 18557-18565.

McGrath, J. P., Jentsch, S. and Varshavsky, A. (1991). "*UBA1*: an essential yeast gene encoding ubiquitin-activating enzyme." Embo J **10**(1): 227-236.

Mizushima, T., Hirao, T., Yoshida, Y., Lee, S. J., Chiba, T., Iwai, K., Yamaguchi, Y., Kato, K., Tsukihara, T. and Tanaka, K. (2004). "Structural basis of sugar-recognising ubiquitin ligase." Nature Structural and Molecular Biology **11**(4): 365-370.

Moore, J. D., Kirk, J. A. and Hunt, T. (2003). "Unmasking the S-Phase-promoting Potential of Cyclin B1." Science **300**: 987-990.

Mullapudi, S., Pullan, L., Bishop, O. T., Khalil, H., Stoops, J. K., Beckmann, R., Kloetzel, P.-M., Krüger, E. and Penczek, P. A. (2004). "Rearrangement of the 16S Precursor Subunits Is Essential for the Formation of the Active 20S Proteasome." Biophys J **87**: 4098-4105.

Muratani, M. and Tansey, W. P. (2003). "How the ubiquitin-proteasome system controls transcription." Nature Reviews Molecular Cell Biology **4**: 1-10.

Murshudov, G. N., Vagin, A. A. and Dodson, E. J. (1997). "Refinement of Macromolecular Structures by the Maximum-Likelihood Method." Acta Crystallogr D Biol Crystallogr **D53**: 240-255.

Musco, G., Stier, G., Kolmerer, B., Adinolfi, S., Martin, S., Frenkiel, T., Gibson, T. and Pastore, A. (2000). "Towards a structural understanding of Friedreich's ataxia: the solution structure of frataxin." Structure **8**: 695-707.

Ohmura, T., Ueda, T., Hashimoto, Y. and Imoto, T. (2001). "Tolerance of point substitution of methionine for isoleucine in hen egg white lysozyme." Protein Engineering **14**(6): 421-425.

Old, R. W. and Primrose, S. B. (1994). Principles of Gene Manipulation, Blackwell Science.

Orlicky, S., Tang, X., Willems, A., Tyers, M. and Sicheri, F. (2003). "Structural basis for phosphodependent substrate selection and orientation by the SCFCdc4 ubiquitin ligase." Cell **112**(2): 243-56.

Otwinowski, Z. (1991). Daresbury Study Weekend proceedings.

Passmore, L. A. and Barford, D. (2004). "Getting into position: the catalytic mechanisms of protein ubiquitylation." Biochem. J. **379**: 513-525.

Pickart, C. M. (2001). "Mechanisms Underlying Ubiquitination." Annu Rev Biochem **70**: 503-533.

Pickart, C. M. and Rose, I. A. (1985). "Functional Heterogeneity of Ubiquitin Carrier Proteins." Journal Biological Chemistry **260**(3): 1573-1581.

Pierce, M. M., Raman, C. S. and Nall, B. T. (1999). "Isothermal Titration Calorimetry of Protein-Protein Interactions." Methods **19**: 213-221.

Pintard, L., Willis, J. H., Willems, A., Johnson, J.-L. F., Srayko, M., Kurz, T., Glaser, S., Mains, P. E., Tyers, M., Bowerman, B. and Peter, M. (2003). "The BTB protein MEL-26 is a substrate-specific adaptor of the CUL-3 ubiquitin-ligase." Nature **425**: 311-316.

Powell, H. R. (1999). "The Rossmann Fourier autoindexing algorithm in MOSFLM." Acta Crystallogr D Biol Crystallogr **D55**: 1690-1695.

Radaev, S. and Sun, P. D. (2002). "Crystallization of protein-protein complexes." Journal of Applied Crystallography **35**: 674-676.

Read, M. A., Brownell, J. E., Gladysheva, T. B., Hottelet, M., Parent, L. A., Coggins, M. B., Pierce, J. W., Podust, V. N., Luo, R.-S., Chau, V. and Palombella, V. J. (2000). "Nedd8 Modification of Cul-1 Activates $SCF^{\beta\text{-}TrCP}$-Dependent Ubiquitination of IκBα." Molecular and Cellular Biology **20**(7): 2326-2333.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular Cloning: a laboratory manual. New York, Cold Spring Harbor Laboratory Press.

Schmidt, M., Zantopf, D., Kraft, R., Kostka, S., Preissner, R. and Kloetzel, P.-M. (1999). "Sequence Information within Proteasomal Prosequences Mediates Efficient Integration of β-subunits into the 20 S Proteasome Complex." Journal Molecular Biology **288**: 117-128.

Schneider, T. R. and Sheldrick, G. M. (2002). "Substructure solution with SHELXD." Acta Crystallogr D Biol Crystallogr **58**(Pt 10 Pt 2): 1772-9.

Schneider, U., Schwenk, H. U. and Bornkamm, G. (1977). "Characterization of EBV-genome negative "null" and "T" cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma." Int J Cancer **19**(5): 621-6.

Schuck, P. (2000). "Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modelling." Biophys J **78**(3): 1606-19.

Schulman, B. A., Carrano, A. C., Jeffrey, P. D., Bowen, Z., Kinnucan, E. R., Finnin, M. S., Elledge, S. J., Harper, J. W., Pagano, M. and Pavletich, N. P. (2000). "Insights into SCF ubiquitin ligases from the structure of the Skp1-Skp2 complex." Nature **408**(6810): 381-6.

Schultz, G. E. and Schirmer, R. H. (1990). Principles of Protein Structure, Springer-Verlag.

Schwartz, D. C. and Hochstrasser, M. (2003). "A superfamily of protein tags: ubiquitin, SUMO and related modifiers." TRENDS in Biochemical Sciences **28**(6): 321-328.

Scott, D. J. and Schuck, P. (2005). A Brief Introduction to the Analytical Ultracentrifugation of Proteins for Beginners.

Shastri, N., Schwab, S. and Serwold, T. (2002). "Producing Nature's Gene Chips: The Generation of Peptides for Display by MHC Class I Molecules." Annu Rev Immunol **20**: 463-493.

Sherr, C. J. (1995). "D-type cyclins." TRENDS in Biochemical Sciences **20**: 187-190.

Siegel, L. M. and Monty, K. J. (1966). "Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulphite and hydroxylamine reductases." Biochim. Biophys. Acta **112**: 346-362.

Sigrist, C. J., Cerutti, L., Hulo, N., Gattiker, A., Falquet, L., Pagni, M., Bairoch, A. and Bucher, P. (2002). "PROSITE: a documented database using patterns and profiles as motif descriptors." Brief Bioinform **3**(3): 265-74.

Skowyra, D., Craig, K. L., Tyers, M., Elledge, S. J. and Harper, J. W. (1997). "F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex." Cell **91**(2): 209-19.

Smith, D. B. and Johnson, K. S. (1988). "Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase." Gene **67**(1): 31-40.

Stebbins, C. E., Kaelin Jr, W. G. and Pavletich, N. P. (1999). "Structure of the VHL-ElonginC-ElonginB Complex: Implication for VHL Tumor Suppressor Function." Science **284**: 455-461.

Stogios, P. J. and Privé, G. G. (2004). "The BACK domain in BTB-kelch proteins." TRENDS in Biochemical Sciences **29**(12): 634-637.

Studier, F. W. and Moffatt, B. A. (1986). "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes." Journal Molecular Biology **189**(1): 113-130.

Sweet, R. M. (1998). "The technology that enables synchrotron structural biology." Nat Struct Biol **5 Suppl**: 654-6.

Tedasco, D., Lukas, J. and Reed, S. I. (2002). "The pRb-related protein p130 is regulated by phosphorylation-dependent proteolysis via a protein- ubiquitin ligase SCF^{Skp2}." Genes Dev **16**: 2946-2957.

Terwilliger, T. C. (2000). "Maximum likelihood density modification." Acta Crystallogr D Biol Crystallogr **D56**: 965-972.

Terwilliger, T. C. and Berendzen, J. (1999). "Automated MAD and MIR structure solution." Acta Crystallogr D Biol Crystallogr **D55**: 849-861.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). "The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." Nucleic Acids Research **24**: 4876-4882.

Tickle, I. J., Laskowski, R. A. and Moss, D. S. (2000). "Rfree and the Rfree ratio. II. Calculation of the expected values and variances of cross-validation statistics in

macromolecular least-squares refinement." Acta Crystallogr D Biol Crystallogr **D56**: 442-450.

Tsou, A., Yang, C.-W., Huang, C.-Y. F., Yu, R. C.-T., Lee, Y.-C. G., Chang, C.-W., Chen, B.-R., Chung, Y.-F., Fann, M.-J., Chi, C.-W., Chiu, J.-H. and Chou, C.-K. (2003). "Identification of a novel cell cycle regulated gene, *HURP*, overexpressed in human hepatocellular carcinoma." Oncogene **22**: 298-307.

Tsvetkov, L. M., Yeh, K.-H., Lee, S.-J., Sun, H. and Zhang, H. (1999). "$p27^{Kip1}$ ubiquitination and degradation is regulated by the SCF^{Skp2} complex through phosphorylated Thr187 in p27." Current Biology **9**: 661-664.

Unno, M., Mizushima, T., Morimoto, Y., Tomisugi, Y., Tanaka, K., Yasuoka, N. and Tsukihara, T. (2002). "The Structure of the Mammalian 20S Proteasome at 2.75 Å Resolution." Structure **10**: 609-618.

Vagin, A. and Teplyakov, A. (1997). "MOLREP: an automated program for molecular replacement." J. Appl. Cryst. **30**: 1022-1025.

Van Nocker, S., Sadis, S., Rubin, D. M., Glickman, M. H., Fu, H., Coux, O., Wefes, I., Finley, D. and Vierstra, R. D. (1996). "The Multiubiquitin-Chain-Binding Protein Mcb1 Is a Component of the 26S Proteasome in *Saccharomyces cerevisiae* and Plays a Nonessential, Substrate-Specific Role in Protein Turnover." Molecular and Cellular Biology **16**(11): 6020-6028.

Vijay-Kumar, S., Bugg, C. E. and Cook, W. J. (1987). "Structure of ubiquitin refined at 1.8 angstroms resolution." Journal Molecular Biology **194**: 531.

Vodermaier, H. V. (2004). "APC/C and SCF: Controlling Each Other and the Cell Cycle." Curr Biol **14**: R787-R796.

von der Lehr, N., Johansson, S., Wu, S., Bahram, F., Castell, A., Cetinkaya, C., Hydbring, P., Weidung, I., Nakayama, K., Nakayama, K. I., Söderberg, O., Kerppola, T. K. and Larsson, L.-G. (2003). "The F-Box Protein Skp2

Participates in c-Myc Proteasomal Degradation and Acts as a Cofactor for c-Myc-Regulated Transcription." Mol Cell **11**: 1189-1200.

Voorhees, P. M., Dees, E. C., O'Neil, B. and Orlowski, R. Z. (2003). "The Proteasome as a Target for Cancer Therapy." Clinical Cancer Research **9**: 6316-6325.

Walden, H., Podgorski, M. S. and Schulman, B. A. (2003). "Insights into the ubiquitin transfer cascade from the structure of the activating enzyme for NEDD8." Nature **422**: 330-334.

Walters, K. J., Kleijnen, M. F., Goh, A. M., Wagner, G. and Howley, P. M. (2002). "Structural Studies of the Interaction between Ubiquitin Family Proteins and Proteasome Subunit S5a." Biochemistry **41**: 1767-1777.

Weissman, A. W. (2001). "Themes and Variations on Ubiquitylation." Nature Reviews Molecular Cell Biology **2**: 169-178.

Wheatley, S. P., Hinchcliffe, E. H., Glotzer, M., Hyman, A. A., Sluder, G. and Wang, Y.-l. (1997). "CDK1 Inactivation Regulates Anaphase Spindle Dynamics and Cytokinesis In Vivo." J. Cell Biol. **138**(2): 385-393.

Whitby, F. G., Masters, E. I., Kramer, L., Knowlton, J. R., Yao, Y., Wang, C. C. and Hill, C. P. (2000). "Structural basis for activation of 20S proteasomes by 11S regulators." Nature **408**: 115-120.

Willems, A. R., Schwab, S. and Tyers, M. (2004). "A hitchhiker's guide to the cullin ubiquitin ligases: SCF and its kin." Biochemica et Biophysica Acta **1695**: 133-170.

Winn, M. D., Isupov, M. N. and Murshudov, G. N. (2001). "Use of TLS parameters to model anisotropic displacements in macromolecular refinement." Acta Crystallogr D Biol Crystallogr **D57**: 122-133.

Winn, M. D., Murshudov, G. N. and Papiz, M. Z. (2003). "Macromolecular TLS Refinement in REFMAC at Moderate Resolutions." Methods in Enzymology **374**: 300-321.

Winston, J. T., Koepp, D. M., Zhu, C., Elledge, S. J. and Harper, J. W. (1999). "A family of mammalian F-box proteins." Curr Biol **9**: 1180-1182.

Wiseman, T., Williston, S., Brandts, J. F. and Lin, L. (1989). "Rapid measurement of binding constants and heats of binding using a new titration calorimeter." Analytical Biochemistry **179**: 131-137.

Wu, G., Xu, G., Schulman, B. A., Jeffrey, P. D., Harper, J. W. and Pavletich, N. P. (2003). "Structure of a beta-TrCP1-Skp1-beta-catenin complex: destruction motif binding and lysine specificity of the SCF(beta-TrCP1) ubiquitin ligase." Mol Cell **11**(6): 1445-56.

Xu, L., Wei, Y., Reboul, J., Vaglio, P., Shin, T.-H., Vidal, M., Elledge, S. J. and Harper, J. W. (2003). "BTB proteins are substrate-specific adaptors in an SCF-like modular ubiquitin ligase containing CUL-3." Nature **425**: 316-321.

Yang, Y., Waters, J. B., Früh, K. and Peterson, P. A. (1992). "Proteasomes are regulated by interferon γ: Implications for antigen processing." Proc Natl Acad Sci U S A **89**: 4928-4932.

Yoshida, Y., Chiba, T., Tokunaga, F., Kawasaki, H., Iwai, K., Suzuki, T., Ito, Y., Matsuoka, K., Yoshida, M., Tanaka, K. and Tai, T. (2002). "E3 ubiquitin ligase that recognizes sugar chains." Nature **418**: 438-442.

Yu, C.-T. R., Hsu, J.-M., Lee, Y.-C. G., Tsou, A.-P., Chou, C.-K. and Huang, C.-Y. F. (2005). "Phosphorylation and Stabilization of HURP by Aurora-A: Implication of HURP as a Transforming Target of Aurora-A." Molecular and Cellular Biology **25**(14): 5789□5800.

Zaiss, D. M. W., Standera, S., Holzhütter, H., Kloetzel, P.-M. and Sijts, A. J. A. M. (1999). "The proteasome inhibitor PI31 competes with PA28 for binding to the 20S proteasome." FEBS Letters **457**: 333-338.

Zaiss, D. M. W., Standera, S., Kloetzel, P.-M. and Sijts, A. J. A. M. (2002). "PI31 is a modulator of proteasome formation and antigen processing." Proc Natl Acad Sci U S A **99**(22): 14344-14349.

Zheng, J., Yang, X., Harrell, J. M., Ryzhikov, S., Shim, E.-H., Lykke-Anderson, K., Wei, N., Sun, H., Kobayashi, R. and Zhang, H. (2002). "CAND1 Binds to Unneddylated CUL1 and Regulates the Formation of SCF Ubiquitin E3 Ligase Complex." Mol Cell **10**: 1519-1526.

Zheng, N., Schulman, B. A., Song, L., Miller, J. J., Jeffrey, P. D., Wang, P., Chu, C., Koepp, D. M., Elledge, S. J., Pagano, M., Conaway, R. C., Conaway, J. W., Harper, J. W. and Pavletich, N. P. (2002). "Structure of the Cul1-Rbx1-Skp1-F box^{Skp2} SCF ubiquitin ligase complex." Nature **416**(6882): 703-9.

Zheng, N., Wang, P., Jeffrey, P. D. and Pavletich, N. P. (2000). "Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin-protein ligases." Cell **102**(4): 533-9.